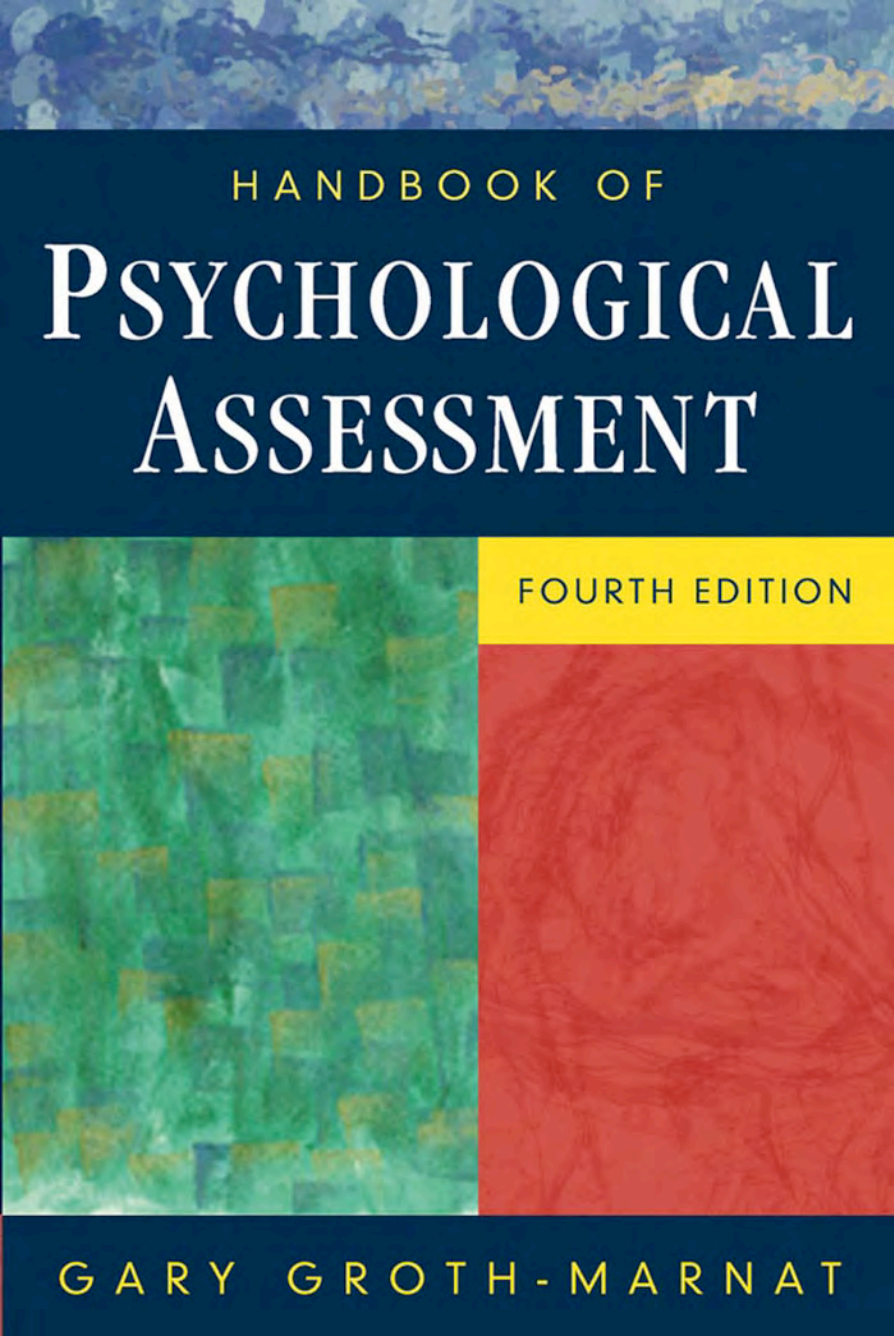

HANDBOOK OF
PSYCHOLOGICAL ASSESSMENT
FOURTH EDITION
GARY GROTH-MARNAT

SHOUT
SHARE
IT

HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

FOURTH EDITION

Gary Groth-Marnat

John Wiley & Sons, Inc.

HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

HANDBOOK OF PSYCHOLOGICAL ASSESSMENT

FOURTH EDITION

Gary Groth-Marnat

John Wiley & Sons, Inc.

MMPI-2™ (Minnesota Multiphasic Personality Inventory-2)™ Test Booklet. Copyright © 1942, 1943 (renewed 1970), 1989 by the Regents of the University of Minnesota. All rights reserved. Used by permission of the University of Minnesota Press. "MMPI-2" and "Minnesota Multiphasic Personality-2" are trademarks owned by the Regents of the University of Minnesota.

This book is printed on acid-free paper. ♾

Copyright © 2003 by John Wiley & Sons, Inc., Hoboken, New Jersey. All rights reserved. Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, e-mail: permcoordinator@wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If legal, accounting, medical, psychological or any other expert assistance is required, the services of a competent professional person should be sought.

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc. is aware of a claim, the product names appear in initial capital or all capital letters. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

For general information on our other products and services please contact our Customer Care Department within the U.S. at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books. For more information about Wiley products, visit our Web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Groth-Marnat, Gary.
Handbook of psychological assessment / Gary Groth-Marnat.—4th ed.
p. cm.
Includes bibliographical references and indexes.
ISBN 0-471-41979-6 (cloth : alk. paper)
1. Psychological tests. 2. Personality assessment. I. Title.
BF176 .G76 2003
150′.28′7—dc21

2002032383

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To Dawn

Contents

Preface **xiii**

Chapter 1 Introduction **1**

Organization of the Handbook 1
Role of the Clinician 3
Patterns of Test Usage in Clinical Assessment 5
Evaluating Psychological Tests 10
Validity in Clinical Practice 22
Clinical Judgment 25
Phases in Clinical Assessment 31
Recommended Reading 36

Chapter 2 Context of Clinical Assessment **37**

Types of Referral Settings 37
Ethical Practice of Assessment 48
Test Bias and Use with Minority Groups 56
Selecting Psychological Tests 62
Computer-Assisted Assessment 66
Recommended Reading 68

Chapter 3 The Assessment Interview **69**

History and Development 70
Issues Related to Reliability and Validity 74
Assets and Limitations 76
The Assessment Interview and Case History 78
Mental Status Examination 84
Interpreting Interview Data 90
Structured Interviews 91
Recommended Reading 101

Chapter 4 Behavioral Assessment 103

History and Development 105
Issues Related to Reliability and Validity 107
Assets and Limitations 110
Strategies of Behavioral Assessment 111
Recommended Reading 128

Chapter 5 Wechsler Intelligence Scales 129

Testing of Intelligence: Pro and Con 129
History and Development 132
Reliability and Validity 134
Assets and Limitations 137
Meaning of IQ Scores 140
Cautions and Guidelines in Administration 142
WAIS-III/WISC-III Successive Level Interpretation Procedure 145
Wechsler Subtests 160
Assessing Brain Damage 178
Assessing Additional Special Populations 186
Short Forms 191
Recommended Reading 195

Chapter 6 Wechsler Memory Scales 197

History and Development 198
Reliability and Validity 201
Assets and Limitations 203
Interpretation Procedure 204
Recommended Reading 212

Chapter 7 Minnesota Multiphasic Personality Inventory 213

History and Development 215
Reliability and Validity 220
Assets and Limitations 223
Administration 230
Interpretation Procedure 231
Computerized Interpretation 239
Validity Scales 240
Clinical Scales 247
Two-Point Codes 270
MMPI-2 Content Scales 294

MMPI-A Content Scales 297
Harris-Lingoes and *Si* Subscales 300
Critical Items 304
MMPI-2 and MMPI-A Supplementary Scales 304
Recommended Reading 308

Chapter 8 Millon Clinical Multiaxial Inventory 311

History and Development 313
Reliability and Validity 317
Assets and Limitations 319
Interpretation Procedure 323
Modifying Indices (Validity Scales) 328
Clinical Personality Patterns 329
Severe Personality Pathology 346
Clinical Syndromes 351
Severe Syndromes 353
Recommended Reading 353

Chapter 9 California Psychological Inventory 355

History and Development 356
Comparison with the MMPI 358
Reliability and Validity 359
Assets and Limitations 360
Interpretation Procedures 363
Vector Scale Interpretation 368
Individual Scales 370
Special Purpose Scales 391
Configural Interpretation 393
Recommended Reading 406

Chapter 10 Rorschach 407

History and Development 408
Reliability and Validity 412
Assets and Limitations 416
Approaching Consensus on Rorschach Validity 419
Administration 420
Scoring 423
Structural Summary 433
Interpretation 440
Recommended Reading 475

Chapter 11 Thematic Apperception Test 477

History and Development 477
Theoretical Perspectives 480
Reliability and Validity 482
Assets and Limitations 485
Administration 488
Typical Themes Elicited 491
Scoring Procedures 505
Interpretation 514
Recommended Reading 515

Chapter 12 Screening and Assessing for Neuropsychological Impairment 517

History and Development 520
Interviewing for Brain Impairment 523
Tests for Screening and Assessing for Neuropsychological Impairment 527
Tests of Visuoconstructive Abilities 528
Mental Activities (Attention and Speed of Information Processing) 548
Memory and Learning 556
Verbal Functions and Academic Skills 565
Tests of Motor Ability 570
Executive Functions 572
Emotional Status and Level of Adjustment 574
Recommended Reading 576

Chapter 13 Brief Instruments for Treatment Planning, Monitoring, and Outcome Assessment 579

Selecting Brief Instruments 580
Symptom Checklist-90-R (SCL-90-R) and Brief Symptom Inventory (BSI) 581
The Beck Depression Inventory 587
State Trait Anxiety Inventory (STAI) 590
Recommended Reading 593

Chapter 14 Psychological Assessment and Treatment Planning 595

Development and Approaches to Treatment Planning 597
A Systematic Approach to Treatment Selection 601
Functional Impairment 603

Social Support 606
Problem Complexity/Chronicity 608
Coping Style 610
Resistance 612
Subjective Distress 614
Problem-Solving Phase 617
Recommended Reading 620

Chapter 15 Psychological Report 621

General Guidelines 623
Format for a Psychological Report 633
Sample Reports 647
Recommended Reading 671

Appendix A 673

Appendix B 677

Appendix C 678

Appendix D 679

Appendix E 680

Appendix F 682

Appendix G 690

Appendix H 691

Appendix I 696

Appendix J 699

Appendix K 700

Appendix L 702

References 703

Author Index 799

Subject Index 811

Preface

Welcome to the fourth edition of *Handbook of Psychological Assessment.* I hope you find this edition to be a clear, useful, and readable guide to conducting psychological assessment. It is readers such as you who have enabled the previous editions to be successful and, because of your interest and feedback, have enabled each edition to be an improvement on the previous ones.

As with the previous editions, I have tried to integrate the best of science with the best of practice. Necessarily, psychological assessment involves technical knowledge. But in presenting this technical knowledge, I have tried to isolate, extract, and summarize in as clear a manner as possible the core information that is required for practitioners to function competently. At the same time, assessment is also about the very human side of understanding, helping, and making decisions about people. I hope I have been able to comfortably blend this technical (science) side with the human. An assessment that does not have at least some heart to it is cold and lacking. To keep in touch with the practitioner/human side of assessment, I have continually maintained an active practice in which I have tried to stay close to and interact with the ongoing personal and professional challenges of practitioners. I hope that within and between the sentences in the book, my active involvement with the world of practice is apparent.

A number of changes in the field of assessment (and psychology in general) are consistent with bringing assessment closer to the person. One is the impact of freedom of information legislation, which means that a report written about a client is more likely to be read by the client; therefore, we as practitioners need to write the report with this in mind. In particular, we must word information about clients in everyday language and in a way that is likely to facilitate personal growth. This is quite consistent with writings by a number of authors who have conceptualized and provided strategies on how to combine assessment with the therapeutic process (therapeutic assessment). This involves not only the use of everyday language, but also a more empathic understanding of the client. It also involves balancing descriptions of clients' weaknesses with their strengths. This is quite consistent with the positive psychology movement that has emerged within mainstream psychology. One of the issues this movement questions is the deeply embedded (medical) model that requires us to identify what is wrong with a person and then go about trying to fix it. Why is this a more effective avenue of change than identifying a client's strengths and then working with the person to enlarge these strengths both as a means in and of itself as well as to overcome any weaknesses? In addition, a client who reads a report describing an endless set of weaknesses will no doubt find it demoralizing (untherapeutic). Unfortunately, *clinical* assessment has still not yet devised a commonly used *multiphasic* instrument of client

strengths. At the same time, I realize that there are certainly a number of referral situations in which capturing this human-centered approach are difficult, such as in forensic contexts when the referral questions may relate to client placement by health professionals or decisions regarding competency.

In addition to this general philosophy of assessment, a number of rather specific developments have been incorporated into the fourth edition (and provide much of the rationale for a further edition). One is the publication of the Wechsler Adult Intelligence Scale-III (WAIS-III; 1997) and the subsequent research on it, which required that I include a WAIS-III supplement as part of a revised third edition in 1999. Readers now find that information in the Wechsler intelligence scales chapter (Chapter 5) itself. A further development has been the publication of and increased popularity of the third edition of the Wechsler Memory Scales (WMS-III). The most recent survey of test use by clinical psychologists ranks it as the ninth most frequently used instrument (and third most popular instrument used by neuropsychologists). At least part of its popularity is the growing importance of assessing memory functions because of an increasing aging population in which distinguishing normal from pathological memory decline has important clinical significance. Other important areas are monitoring the effects of medication to improve memory; detecting cognitive decline resulting from substance abuse; and detecting impairment caused by neurotoxic exposure or the impact of brain trauma, stroke, or the progression of brain disease (Alzheimer's disease, AIDS-related dementia). As a result, a brief introductory chapter (Chapter 6) was developed on the Wechsler memory scales.

A further change is the inclusion of a chapter on brief instruments for treatment planning, monitoring, and outcome assessment (Chapter 13). This chapter was considered essential because of the increasing emphasis on providing empirical support for the effectiveness of clinical interventions. Many managed care organizations either encourage or require such accountability. It is hoped that this chapter provides readers with a preliminary working knowledge of the three instruments used most frequently in this process (Symptom Checklist 90-R, Beck Depression Inventory, State Trait Anxiety Inventory). Because of the decreasing use of projective drawings combined with continued research that questions the validity of many, if not most, of the interpretations based on projective drawing data, the chapter on projective drawings included in the previous three editions was omitted to make room for these new chapters (Chapters 6 and 13).

The field of psychological assessment is continually expanding and evolving. Sometimes it is difficult to keep up with the sheer number of publications. Much of this is reflected not in separate chapters but in numerous small updates and changes within chapters. For example, there have been new surveys in test practice and publication of new ethical standards, structured clinical interviews have changed to keep pace with the *DSM-IV,* the MMPI-2 has altered its profiles to include newer (particularly validity) scales, considerable (heated) debate has revolved around the Rorschach, a new edition of the Comprehensive System has been released, new approaches and theories have been used with the TAT, and refinements have occurred in treatment planning. In addition, the continuous publication of new books and research further refines the practice of assessment. As with previous editions, my goal has been to display the utmost in conscientiousness but fall just short of obsessiveness.

Writing this handbook has required a huge effort. Major thanks go to my students, who continually keep me in line with what works as well as what doesn't work. The refinements in the handbook reflect their thoughts and feedback through the process of continually "test driving" new methods on them (or, more collaboratively, taking the journey with them as they learn the skills of assessment). A number of students have been particularly helpful, including Jennifer Crowhurst, Brendan Dellar, Kim Estep, Melinda Jeffs, Gemma Johnston, and Julie Robarts. Dawn Erickson has also been particularly helpful with both support and last-minute work on the references (see dedication for further details). Further thanks go to Greg Meyer, Howard Garb, and John Exner for their perspectives on the Rorschach controversy and their advice and willingness for me to reproduce materials. Special thanks go to Larry Beutler and the Counseling/Clinical/School Program at the University of California, Santa Barbara, who allowed me to be a visiting scholar with their program during 2001/2002. The use of office space, library facilities, and students was minor (but greatly appreciated) when compared with the wealth of ideas, humor, and opportunities for co-teaching, colleagueship, and friendship. Long-term thanks also go to Dorothy (Gita) Morena, who began as a co-author on the first edition more than 20 years ago. As always, the team at John Wiley & Sons has been a pleasure to work with. In particular, Jennifer Simon has been instrumental in humoring, inspiring, and cajoling this fourth edition into existence ("When is that manuscript *really* going to be delivered?"). Pam Blackmon and Nancy Land at Publications Development Company of Texas have also done a fantastic job of turning the raw manuscript into pages, ink, and binding. Finally, Australia in general and Curtin University in particular have been a fine home and the place where both my career and the *Handbook of Psychological Assessment* (all three editions) have been nurtured and developed. My thanks to all the staff, friends, and colleagues who supported me and inspired me to make this happen. Having now moved back to the United States (after 18 years), I have left a big part of myself there and brought a big part of Australia back with me. I owe much of my career (which has been greatly guided by and dominated by the *Handbook of Psychological Assessment*) to you.

Chapter 1

INTRODUCTION

The *Handbook of Psychological Assessment* is designed to develop a high level of practitioner competence by providing relevant practical, as well as some theoretical, material. It can serve as both a reference and instructional guide. As a reference book, it aids in test selection and the development of a large number and variety of interpretive hypotheses. As an instructional text, it provides students with the basic tools for conducting an integrated psychological assessment. The significant and overriding emphasis in this book is on assessing areas that are of practical use in evaluating individuals in a clinical context. It is applied in its orientation, and for the most part, I have kept theoretical discussions to a minimum. Many books written on psychological testing and the courses organized around these books focus primarily on test theory, with a brief overview of a large number of tests. In contrast, my intent is to focus on the actual processes that practitioners go through during assessment. I begin with such issues as role clarification and evaluation of the referral question and end with treatment planning and the actual preparation of the report itself. Although I have included some material on test theory, my purpose is to review those areas that are most relevant in evaluating tests before including them in a battery.

One of the crucial skills that I hope readers of this text will develop, or at least have enhanced, is a realistic appreciation of the assets and limitations of assessment. This includes an appraisal of psychological assessment as a general strategy as well as an awareness of the assets and limitations of specific instruments and procedures. A primary limitation of assessment lies in the incorrect handling of the data, which is not integrated in the context of other sources of information (behavioral observations, history, other test scores). Also, the results are not presented in a way that helps solve the unique problems clients or referral sources are confronting. To counter these limitations, the text continually provides practitioners with guidelines for integrating and presenting the data in as useful a manner as possible. The text is thus not so much a book on test interpretation (although this is an important component) but on test integration within the wider context of assessment. As a result, psychologists should be able to create reports that are accurate, effective, concise, and highly valued by the persons who receive them.

ORGANIZATION OF THE HANDBOOK

My central organizational plan for the *Handbook of Psychological Assessment* replicates the sequence practitioners follow when performing an evaluation. They are initially

concerned with clarifying their roles, ensuring that they understand all the implications of the referral question, deciding which procedures would be most appropriate for the assessment, and reminding themselves of the potential problems associated with clinical judgment (Chapter 1). They also need to understand the context in which they will conduct the assessment. This understanding includes appreciating the issues, concerns, terminology, and likely roles of the persons from these contexts. Practitioners also must have clear ethical guidelines, know how to work with persons from diverse backgrounds, and recognize issues related to computer-assisted assessment and the ways that the preceding factors might influence their selection of procedures (see Chapter 2).

Once practitioners have fully understood the preliminary issues discussed in Chapters 1 and 2, they must select different strategies of assessment. The three major strategies are interviewing, observing behavior, and psychological testing. An interview is likely to occur during the initial phases of assessment and is also essential in interpreting test scores and understanding behavioral observations (see Chapter 3). The assessment of actual behaviors might also be undertaken (see Chapter 4). Behavioral assessment might be either an end in itself, or an adjunct to testing. It might involve a variety of strategies such as the measurement of overt behaviors, cognitions, alterations in physiology, or relevant measures from self-report inventories.

The middle part of the book (Chapters 5 through 13) provides a general overview of the most frequently used tests. Each chapter begins with an introduction to the test in the form of a discussion of its history and development, current evaluation, and procedures for administration. The main portions of these chapters provide a guide for interpretation, which includes such areas as the meaning of different scales, significant relations between scales, frequent trends, and the meaning of unusually high or low scores. When appropriate, there are additional subsections. For example, Chapter 5, "Wechsler Intelligence Scales," includes additional sections on the meaning of IQ scores, estimating premorbid IQ, and assessing special populations. Likewise, Chapter 11, "Thematic Apperception Test," includes a summary of Murray's theory of personality because a knowledge of his concepts is a prerequisite for understanding and interpreting the test. Chapter 12, "Screening and Assessing for Neuropsychological Impairment," varies somewhat from the preceding format in that it is more a compendium and interpretive guide to some of the most frequently used short neuropsychological tests, along with a section on the special considerations in conducting a neuropsychological interview. This organization reflects the current emphasis on and strategies for assessing patients with possible neuropsychological dysfunction.

Several of the chapters on psychological tests are quite long, particularly those for the Wechsler scales, Minnesota Multiphasic Personality Inventory, and neuropsychological screening and assessment. These chapters include extensive summaries of a wide variety of interpretive hypotheses intended for reference purposes when practitioners must generate interpretive hypotheses based on specific test scores. To gain initial familiarity with the tests, I recommend that practitioners or students carefully read the initial sections (history and development, psychometric properties, etc.) and then skim through the interpretation sections more quickly. This provides the reader with a basic familiarity with the procedures and types of data obtainable from the tests. As practical test work progresses, clinicians can then study the interpretive hypotheses in greater depth and gradually develop more extensive knowledge of the scales and their interpretation.

Based primarily on current frequency of use, the following tests are covered in this text: the Wechsler intelligence scales (WAIS-III/WISC-III), Wechsler Memory Scales (WMS-III), Minnesota Multiphasic Personality Inventory (MMPI-2), Millon Clinical Multiaxial Inventory (MCMI-III), Bender Visual Motor Gestalt Test (along with other frequently used neuropsychological tests), Rorschach, and the Thematic Apperception Test (TAT; Camara, Nathan, & Puente, 2000; C. Piotrowski & Zalewski, 1993; Watkins, 1991; Watkins, Campbell, Nieberding, & Hallmark, 1995). The California Personality Inventory (CPI) was selected because of the importance of including a broad-based inventory of normal functioning along with its excellent technical development and relatively large research base (Anastasi & Urbina, 1997; Baucom, 1985; Gough, 2000; Wetzler, 1990). I also included a chapter on the most frequently used brief, symptom-focused inventories because of the increasing importance of monitoring treatment progress and outcome in a cost- and time-efficient managed care environment (Eisman, 2000; C. Piotrowski, 1999). The preceding instruments represent the core assessment devices used by most practitioners.

Finally, the clinician must generate relevant treatment recommendations and integrate the assessment results into a psychological report. Chapter 14 provides a systematic approach for working with assessment results to develop practical, empirically supported treatment recommendations. Chapter 15 presents guidelines for report writing, a report format, and four sample reports representative of the four most common types of referral settings (medical setting, legal context, educational context, psychological clinic). Thus, the chapters follow a logical sequence and provide useful, concise, and practical knowledge.

ROLE OF THE CLINICIAN

The central role of clinicians conducting assessments should be to answer specific questions and aid in making relevant decisions. To fulfill this role, clinicians must integrate a wide range of data and bring into focus diverse areas of knowledge. Thus, they are not merely administering and scoring tests. A useful distinction to highlight this point is the contrast between a psychometrist and a clinician conducting psychological assessment (Maloney & Ward, 1976; Matarazzo, 1990). Psychometrists tend to use tests merely to obtain data, and their task is often perceived as emphasizing the clerical and technical aspects of testing. Their approach is primarily data oriented, and the end product is often a series of traits or ability descriptions. These descriptions are typically unrelated to the person's overall context and do not address unique problems the person may be facing. In contrast, psychological assessment attempts to evaluate an individual in a problem situation so that the information derived from the assessment can somehow help with the problem. Tests are only one method of gathering data, and the test scores are not end products, but merely means of generating hypotheses. Psychological assessment, then, places data in a wide perspective, with its focus being problem solving and decision making.

The distinction between psychometric testing and psychological assessment can be better understood and the ideal role of the clinician more clearly defined by briefly elaborating on the historical and methodological reasons for the development of the psychometric approach. When psychological tests were originally developed, group

measurements of intelligence met with early and noteworthy success, especially in military and industrial settings where individual interviewing and case histories were too expensive and time consuming. An advantage of the data-oriented intelligence tests was that they appeared to be objective, which would reduce possible interviewer bias. More important, they were quite successful in producing a relatively high number of true positives when used for classification purposes. Their predictions were generally accurate and usable. However, this created the early expectation that all assessments could be performed using the same method and would provide a similar level of accuracy and usefulness. Later assessment strategies often tried to imitate the methods of earlier intelligence tests for variables such as personality and psychiatric diagnosis.

A further development consistent with the psychometric approach was the strategy of using a "test battery." It was reasoned that if a single test could produce accurate descriptions of an ability or trait, administering a series of tests could create a total picture of the person. The goal, then, was to develop a global, yet definitive, description for the person using purely objective methods. This goal encouraged the idea that the tool (psychological test) was the best process for achieving the goal, rather than being merely one technique in the overall assessment procedure. Behind this approach were the concepts of *individual differences* and *trait psychology.* These assume that one of the best ways to describe the differences among individuals is to measure their strengths and weaknesses with respect to various traits. Thus, the clearest approach to the study of personality involved developing a relevant taxonomy of traits and then creating tests to measure these traits. Again, there was an emphasis on the tools as primary, with a de-emphasis on the input of the clinician. These trends created a bias toward administration and clerical skills. In this context, the psychometrist requires little, if any, clinical expertise other than administering, scoring, and interpreting tests. According to such a view, the most preferred tests would be machine-scored true-false or multiple choice-constructed so that the normed scores, rather than the psychometrist, provide the interpretation.

The objective psychometric approach is most appropriately applicable to ability tests such as those measuring intelligence or mechanical skills. Its usefulness decreases, however, when users attempt to assess personality traits such as dependence, authoritarianism, or anxiety. Personality variables are far more complex and, therefore, need to be validated in the context of history, behavioral observations, and interpersonal relationships. For example, a *T* score of 70 on the MMPI-2 scale 9 (mania) takes on an entirely different meaning for a high-functioning physician than for an individual with a poor history of work and interpersonal relationships. When the purely objective psychometric approach is used for the evaluation of problems in living (neurosis, psychosis, etc.), its usefulness is questionable.

Psychological assessment is most useful in the understanding and evaluation of personality and especially of problems in living. These issues involve a particular problem situation having to do with a specific individual. The central role of the clinician performing psychological assessment is that of an expert in human behavior who must deal with complex processes and understand test scores in the context of a person's life. The clinician must have knowledge concerning problem areas and, on the basis of this knowledge, form a general idea regarding behaviors to observe and areas in which to collect relevant data. This involves an awareness and appreciation of multiple causation, interactional influences, and multiple relationships. As Woody (1980) has stated, "Clinical

assessment is individually oriented, but it always considers social existence; the objective is usually to help the person solve problems."

In addition to an awareness of the role suggested by psychological assessment, clinicians should be familiar with core knowledge related to measurement and clinical practice. This includes descriptive statistics, reliability (and measurement error), validity (and the meaning of test scores), normative interpretation, selection of appropriate tests, administration procedures, variables related to diversity (ethnicity, race, age, gender), testing individuals with disabilities, and an appropriate amount of supervised experience (Turner, DeMers, Fox, & Reed, 2001). Persons performing psychological assessment should also have basic knowledge related to the demands, types of referral questions, and expectations of various contexts—particularly employment, education, vocational/career, health care (psychological, psychiatric, medical), and forensic. Furthermore, clinicians should know the main interpretive hypotheses in psychological testing and be able to identify, sift through, and evaluate a series of hypotheses to determine which are most relevant and accurate. For each assessment device, clinicians must understand conceptually what they are trying to test. Thus, rather than merely knowing the labels and definitions for various types of anxiety or thought disorders, clinicians should also have in-depth operational criteria for them. For example, the concept of intelligence, as represented by the IQ score, can sometimes appear misleadingly straightforward. Intelligence test scores can be complex, however, involving a variety of cognitive abilities, the influence of cultural factors, varying performance under different conditions, and issues related to the nature of intelligence. Unless clinicians are familiar with these areas, they are not adequately prepared to handle IQ data.

The above knowledge should be integrated with relevant general coursework, including abnormal psychology, the psychology of adjustment, clinical neuropsychology, psychotherapy, and basic case management. A problem in many training programs is that, although students frequently have a knowledge of abnormal psychology, personality theory, and test construction, they usually have insufficient training to integrate their knowledge into the interpretation of test results. Their training focuses on developing competency in administration and scoring, rather than on knowledge relating to what they are testing.

The approach in this book is consistent with that of psychological assessment: Clinicians should be not only knowledgeable about traditional content areas in psychology and the various contexts of assessment, but also able to integrate the test data into a relevant description of the person. This description, although focusing on the individual, should take into account the complexity of his or her social environment, personal history, and behavioral observations. Yet, the goal is not merely to describe the person, but rather to develop relevant answers to specific questions, aid in problem solving, and facilitate decision making.

PATTERNS OF TEST USAGE IN CLINICAL ASSESSMENT

Psychological assessment is crucial to the definition, training, and practice of professional psychology. Fully 91% of all practicing psychologists engage in assessment (Watkins et al., 1995), and 64% of all nonacademic advertisements listed assessment as

an important prerequisite (Kinder, 1994). Assessment skills are also strong prerequisites for internships and postdoctoral training. The theory and instruments of assessment can be considered the very foundation of clinical investigation, applied research, and program evaluation. In many ways, psychological assessment is professional psychology's unique contribution to the wider arena of clinical practice. The early professional psychologists even defined themselves largely in the context of their role as psychological testers. Practicing psychologists currently spend 10% to 25% of their time conducting psychological assessment (Camara et al., 2000; Watkins, 1991; Watkins et al., 1995).

Although assessment has always been a core, defining feature of professional psychology, the patterns of use and relative importance of assessment have changed with time. During the 1940s and 1950s, psychological testing was frequently the single most important activity of professional psychologists. In contrast, the past 50 years have seen psychologists become involved in a far wider diversity of activities. Lubin and his colleagues (Lubin, Larsen, & Matarazzo, 1984; Lubin, Larsen, Matarazzo, & Seever, 1985, 1986) found that the average time spent performing assessment across five treatment settings was 44% in 1959, 29% in 1969, and only 22% in 1982. The average time spent in 1982 performing assessments in the five different settings ranged from 14% in counseling centers to 31% in psychiatric hospitals (Lubin et al., 1984, 1985, 1986). A recent survey found that the vast majority (81%) spend 0 to 4 hours a week, 15% spend 5 to 20 hours a week, and 4% spend more than 20 hours a week conducting assessments (Camara et al., 2000). The gradual decrease in the total time spent in assessment is due in part to the widening role of psychologists. Whereas in the 1940s and 1950s a practicing psychologist was almost synonymous with a tester, professional psychologists currently are increasingly involved in administration, consultation, organizational development, and many areas of direct treatment (Bamgbose, Smith, Jesse, & Groth-Marnat, 1980; Groth-Marnat, 1988; Groth-Marnat & Edkins, 1996). Decline in testing has also been attributed to disillusionment with the testing process based on criticisms about the reliability and validity of many assessment devices (Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Wood, Lilienfeld, Garb, & Nezworski, 2000; Ziskin & Faust, 1995). Testing activity has also decreased because of reductions in reimbursements from managed care (C. Piotrowski, 1999). In addition, psychological assessment has come to include a wide variety of activities beyond merely the administration and interpretation of traditional tests. These include conducting structured and unstructured interviews, behavioral observations in natural settings, observations of interpersonal interactions, neuropsychological assessment, and behavioral assessment.

The relative popularity of different traditional psychological tests has been surveyed since 1935 in many settings such as academic institutions, psychiatric hospitals, counseling centers, veterans administration centers, institutions for the developmentally disabled, private practice, and various memberships and professional organizations. Surveys of test usage have usually found that the 10 most frequently used tests are the Wechsler intelligence scales, Minnesota Multiphasic Personality Inventory, Rorschach, Bender Visual Motor Gestalt Test, Thematic Apperception Test, projective drawings (Human Figure Drawing, House-Tree-Person), Wechsler Memory Scale, Beck Depression Inventory, Millon Clinical Multiaxial Inventories, and California Psychological Inventory (Camara et al., 2000; Kamphaus, Petoskey, & Rowe, 2000; Lubin et al., 1985; C. Piotrowski & Zalewski, 1993; Watkins, 1991; Watkins et al.,

1995). The pattern for the 10 most popular tests has remained quite stable since 1969 except that the Millon Clinical Multiaxial Inventory is now ranked number 10 and Human Figure Drawings have decreased to 13 (Camara et al., 2000). The pattern of test usage varies somewhat across different studies and varies considerably from setting to setting. Schools and centers for the intellectually disabled emphasize tests of intellectual abilities such as the WISC-III; counseling centers are more likely to use vocational interest inventories; and psychiatric settings emphasize tests assessing level of pathology such as the MMPI-2 or MCMI-III.

One clear change in testing practices has been a relative decrease in the use and status of projective techniques (Groth-Marnat, 2000b; C. Piotrowski, 1999). Criticisms have been wide ranging but have centered on overly complex scoring systems, questionable norms, subjectivity of scoring, poor predictive utility, and inadequate or even nonexistent validity (Garb et al., 2001; Pruitt, Smith, Thelen, & Lubin, 1985; D. Smith & Dumont, 1995; Wood, Lilienfeld, Nexworski, & Garb, 2000). Further criticisms include the extensive time required to effectively learn the techniques, heavy reliance of projective techniques on psychoanalytic theory, and the greater time and cost efficiency of alternative objective tests. These criticisms have usually occurred from within the academic community where they are used less and less for research purposes (C. Piotrowski, 1999; C. Piotrowski & Zalewski, 1993; Watkins, 1991). As a result of these criticisms, there has been a slight but still noteworthy reduction in the use of the standard projective tests in professional practice (Camara et al., 2000; Kamphaus et al., 2000; C. Piotrowski, 1999). Although there has been a reduction, the Rorschach and TAT are still among the ten most frequently used instruments in adult clinical settings. This can be attributed to lack of time available for practitioners to learn new techniques, expectations that students in internships know how to use them (C. Piotrowski & Zalewski, 1993), unavailability of other practical alternatives, and the fact that clinical experience is usually given more weight by practitioners than empirical evidence (Beutler, Williams, Wakefield, & Entwistle, 1995). This suggests distance between the quantitative, theoretical world of the academic and the practical, problem-oriented world of the practitioner. In fact, assessment practices in many professional settings seem to have little relationship to the number of research studies done on assessment tools, attitudes by academic faculty (C. Piotrowski & Zalewski, 1993), or the psychometric quality of the test. In contrast to the continued use of projective instruments in adult clinical settings, psychologists in child settings have largely supplanted projective instruments with behavior rating scales such as the Behavior Assessment System for Children, Connor's Parent/Teacher Rating Scale, and the Achenbach Child Behavior Checklist (Kamphaus et al., 2000).

The earliest form of assessment was through clinical interview. Clinicians such as Freud, Jung, and Adler used unstructured interaction to obtain information regarding history, diagnosis, or underlying structure of personality. Later clinicians taught interviewing by providing outlines of the areas that should be discussed. During the 1960s and 1970s, much criticism was directed toward the interview, leading many psychologists to perceive interviews as unreliable and lacking empirical validation. Tests, in many ways, were designed to counter the subjectivity and bias of interview techniques. During the 1980s and 1990s, a wide variety of structured interview techniques gained popularity and have often been found to be reliable and valid indicators of a client's

level of functioning. Structured interviews such as the Diagnostic Interview Schedule (DIS; Robins, Helzer, Cottler, & Goldring, 1989), Structured Clinical Interview for the *DSM* (SCID; Spitzer, Williams, & Gibbon, 1987), and Renard Diagnostic Interview (Helzer, Robins, Croughan, & Welner, 1981) are often given preference over psychological tests. These interviews, however, are very different from the traditional unstructured approaches. They have the advantage of being psychometrically sound even though they might lack important elements of rapport, idiographic richness, and flexibility that characterize less structured interactions.

A further trend has been the development of neuropsychological assessment (see Groth-Marnat, 2000a). The discipline is a synthesis between behavioral neurology and psychometrics and was created from a need to answer questions such as the nature of a person's organic deficits, severity of deficits, localization, and differentiating between functional versus organic impairment. The pathognomonic sign approach and the psychometric approaches are two clear traditions that have developed in the discipline. Clinicians relying primarily on a pathognomonic sign approach are more likely to interpret specific behaviors such as perseverations or weaknesses on one side of the body, which are highly indicative of the presence and nature of organic impairments. These clinicians tend to rely on the tradition of assessment associated with Luria (Bauer, 1995; Luria, 1973) and base their interview design and tests on a flexible method of testing possible hypotheses for different types of impairment. In contrast, the more quantitative tradition represented by Reitan and his colleagues (Reitan & Wolfson, 1993; Russell, 2000) is more likely to rely on critical cutoff scores, which distinguish between normal and brain-damaged persons. Reitan and Wolfson (1985, 1993) have recommended using an *impairment index,* which is the proportion of brain-sensitive tests that fall into the brain-damaged range. In actual practice, most clinical neuropsychologists are more likely to combine the psychometric and pathognomonic sign approaches. The two major neuropsychological test batteries currently used in the United States are the Luria-Nebraska Neuropsychological Battery (Golden, Purisch, & Hammeke, 1985) and the Halstead Reitan Neuropsychological Test Battery (Reitan & Wolfson, 1993). A typical neuropsychological battery might include tests specifically designed to assess organic impairment along with tests such as the MMPI, Wechsler intelligence scales, and the Wide Range Achievement Test (WRAT-III). As a result, extensive research over the past 10 to 15 years has been directed toward developing a greater understanding of how the older and more traditional tests relate to different types and levels of cerebral dysfunction.

During the 1960s and 1970s, behavior therapy was increasingly used and accepted. Initially, behavior therapists were concerned with an idiographic approach to the functional analysis of behavior. As their techniques became more sophisticated, formalized methods of behavioral assessment began to arise. These techniques arose in part from dissatisfaction with the *Diagnostic and Statistical Manual of Mental Disorders, 2nd Edition* (*DSM-II*; American Psychiatric Association, 1968) methods of diagnosis as well as from a need to have assessment relate more directly to treatment and its outcomes. There was also a desire to be more accountable for documenting behavior change over time. For example, if behaviors related to anxiety decreased after therapy, the therapist should be able to demonstrate that the treatment had been successful. Behavioral assessment could involve measurements of movements (behavioral checklists,

behavioral analysis), physiological responses (Galvanic Skin Response [GSR], Electromyograph [EMG]) or self-reports (self-monitoring, Beck Depression Inventory, assertiveness scales). Whereas the early behavioral assessment techniques showed little concern with the psychometric properties of their instruments, there has been an increasing push to have them meet adequate levels of reliability and validity (First, Frances, Widiger, Pincus, & Davis, 1992; Follette & Hayes, 1992). Despite the many formalized techniques of behavioral assessment, many behavior therapists feel that an unstructured idiographic approach is most appropriate.

Traditional means of assessment, then, have decreased because of an overall increase in other activities of psychologists and an expansion in the definition of assessment. Currently, a psychologist doing assessment might include such techniques as interviewing, administering, and interpreting traditional psychological tests (MMPI-2/MMPI-A, WAIS-III, etc.), naturalistic observations, neuropsychological assessment, and behavioral assessment. In addition, professional psychologists might be required to assess areas that were not given much emphasis before the 1980s—personality disorders (borderline personality, narcissism), stress and coping (life changes, burnout, existing coping resources), hypnotic responsiveness, psychological health, adaptation to new cultures, and the changes associated with increasing modernization. Additional areas might include family systems interactions, relation between a person and his or her environment (social climate, social supports), cognitive processes related to behavior disorders, and level of personal control (self-efficacy). All these require clinicians to be continually aware of new and more specific assessment devices and to maintain flexibility in the approaches they take.

The future of psychological assessment will probably be most influenced by the trends toward computerized assessment, adaptation to managed health care, and distance health care delivery (Groth-Marnat, 2000b). Computerized assessment is likely to enhance efficiency through rapid scoring, complex decision rules, reduction in client-practitioner contact, novel presentation of stimuli (i.e., virtual reality), and generation of interpretive hypothesis. Future assessments are also likely to tailor the presentation of items based on the client's previous responses. Unnecessary items will not be given with one result being that a larger amount of information will be obtained through the presentation of relatively fewer items. This time efficiency is in part stimulated by the cost savings policies of managed care, which require psychologists to demonstrate the cost-effectiveness of their services (Groth-Marnat, 1999; Groth-Marnat & Edkins, 1996). In assessment, this means linking assessment with treatment planning. Thus, psychological reports of the future are likely to spend relatively less time on client dynamics and more time on details related to specific intervention strategies. Whereas considerable evidence supports the cost-effectiveness of using psychological tests in organizational contexts, health care similarly needs to demonstrate that assessment can increase the speed of treatment as well as optimize treatment outcome (see Groth-Marnat, 1999).

A further challenge and area for development is the role assessment will play in distance health (Leigh & Zaylor, 2000; M. A. Smith & Senior, 2001). It might be particularly important for users of these facilities to be screened (or screen themselves) in order to optimally tailor interventions. In addition, distance assessment as a means in and of itself is likely to become important as well. This might require professional psychologists to change their traditional face-to-face role to one of developing and

monitoring new applications as well as consulting/collaborating with clients regarding the results of assessments derived from the computer.

EVALUATING PSYCHOLOGICAL TESTS

Before using a psychological test, clinicians should investigate and understand the theoretical orientation of the test, practical considerations, the appropriateness of the standardization sample, and the adequacy of its reliability and validity. Often, helpful descriptions and reviews that relate to these issues can be found in past and future editions of the *Mental Measurements Yearbook* (Impara & Plake, 1998), *Tests in Print* (L. Murphy, Impara, & Plake, 1999), *Tests: A Comprehensive Reference for Assessment in Psychology, Education, and Business* (Maddox, 1997), and *Measures for Clinical Practice: A Sourcebook* (Corcoran, 2000). Reviews can also be found in assessment-related journals such as the *Journal of Personality Assessment, Journal of Psychoeducational Assessment,* and *Educational and Psychological Measurement.* Test users should carefully review the manual accompanying the test. Table 1.1 outlines the more important questions that should be answered. The issues outlined in this table are discussed further. The discussion reflects the practical orientation of this text by focusing on

Table 1.1 Evaluating a psychological test

Theoretical Orientation

1. Do you adequately understand the theoretical construct the test is supposed to be measuring?
2. Do the test items correspond to the theoretical description of the construct?

Practical Considerations

1. If reading is required by the examinee, does his or her ability match the level required by the test?
2. How appropriate is the length of the test?

Standardization

1. Is the population to be tested similar to the population the test was standardized on?
2. Was the size of the standardization sample adequate?
3. Have specialized subgroup norms been established?
4. How adequately do the instructions permit standardized administration?

Reliability

1. Are reliability estimates sufficiently high (generally around .90 for clinical decision making and around .70 for research purposes)?
2. What implications do the relative stability of the trait, the method of estimating reliability, and the test format have on reliability?

Validity

1. What criteria and procedures were used to validate the test?
2. Will the test produce accurate measurements in the context and for the purpose for which you would like to use it?

problems that clinicians using psychological tests are likely to confront. It is not intended to provide a comprehensive coverage of test theory and construction; if a more detailed treatment is required, the reader is referred to one of the many texts on psychological testing (e.g., Anastasi & Urbina, 1997; R. Kaplan & Saccuzzo, 2001).

Theoretical Orientation

Before clinicians can effectively evaluate whether a test is appropriate, they must understand its theoretical orientation. Clinicians should research the construct that the test is supposed to measure and then examine how the test approaches this construct (see S. Haynes, Richard, & Kubany, 1995). This information can usually be found in the test manual. If for any reason the information in the manual is insufficient, clinicians should seek it elsewhere. Clinicians can frequently obtain useful information regarding the construct being measured by carefully studying the individual test items. Usually the manual provides an individual analysis of the items, which can help the potential test user evaluate whether they are relevant to the trait being measured.

Practical Considerations

A number of practical issues relate more to the context and manner in which the test is used than to its construction. First, tests vary in terms of the level of education (especially reading skills) that examinees must have to understand them adequately. The examinee must be able to read, comprehend, and respond appropriately to the test. Second, some tests are too long, which can lead to a loss of rapport with, or extensive frustration on the part of, the examinee. Administering short forms of the test may reduce these problems, provided these forms have been properly developed and are treated with appropriate caution. Finally, clinicians have to assess the extent to which they need training to administer and interpret the instrument. If further training is necessary, a plan must be developed for acquiring this training.

Standardization

Another central issue relates to the adequacy of norms (see Cicchetti, 1994). Each test has norms that reflect the distribution of scores by a standardization sample. The basis on which individual test scores have meaning relates directly to the similarity between the individual being tested and the sample. If a similarity exists between the group or individual being tested and the standardization sample, adequate comparisons can be made. For example, if the test was standardized on college students between the ages of 18 and 22, useful comparisons can be made for college students in that age bracket (if we assume that the test is otherwise sufficiently reliable and valid). The more dissimilar the person is from this standardization group (e.g., over 70 years of age with low educational achievement), the less useful the test is for evaluation. The examiner may need to consult the literature to determine whether research that followed the publication of the test manual has developed norms for different groups. This is particularly important for tests such as the MMPI and the Rorschach in which norms for younger populations have been published.

Three major questions that relate to the adequacy of norms must be answered. The first is whether the standardization group is representative of the population on which the examiner would like to use the test. The test manual should include sufficient information to determine the representativeness of the standardization sample. If this information is insufficient or in any way incomplete, it greatly reduces the degree of confidence with which clinicians can use the test. The ideal and current practice is to use stratified random sampling. However, because this can be an extremely costly and time-consuming procedure, many tests are quite deficient in this respect. The second question is whether the standardization group is large enough. If the group is too small, the results may not give stable estimates because of too much random fluctuation. Finally, a good test has specialized subgroup norms as well as broad national norms. Knowledge relating to subgroup norms gives examiners greater flexibility and confidence if they are using the test with similar subgroup populations (see Dana, 2000). This is particularly important when subgroups produce sets of scores that are significantly different from the normal standardization group. These subgroups can be based on factors such as ethnicity, sex, geographic location, age, level of education, socioeconomic status, or urban versus rural environment. Knowledge of each of these subgroup norms allows for a more appropriate and meaningful interpretation of scores.

Standardization can also refer to administration procedures. A well-constructed test should have instructions that permit the examiner to give the test in a structured manner similar to that of other examiners and also to maintain this standardized administration between one testing session and the next. Research has demonstrated that varying the instructions between one administration and the next can alter the types and quality of responses the examinee makes, thereby compromising the test's reliability. Standardization of administration should refer not only to the instructions, but also to ensuring adequate lighting, quiet, no interruptions, and good rapport.

Reliability

The reliability of a test refers to its degree of stability, consistency, predictability, and accuracy. It addresses the extent to which scores obtained by a person are the same if the person is reexamined by the same test on different occasions. Underlying the concept of reliability is the possible range of error, or error of measurement, of a single score. This is an estimate of the range of possible random fluctuation that can be expected in an individual's score. It should be stressed, however, that a certain degree of error or noise is always present in the system, from such factors as a misreading of the items, poor administration procedures, or the changing mood of the client. If there is a large degree of random fluctuation, the examiner cannot place a great deal of confidence in an individual's scores. The goal of a test constructor is to reduce, as much as possible, the degree of measurement error, or random fluctuation. If this is achieved, the difference between one score and another for a measured characteristic is more likely to result from some true difference than from some chance fluctuation.

Two main issues relate to the degree of error in a test. The first is the inevitable, natural variation in human performance. Usually the variability is less for measurements of ability than for those of personality. Whereas ability variables (intelligence, mechanical aptitude, etc.) show gradual changes resulting from growth and development,

many personality traits are much more highly dependent on factors such as mood. This is particularly true in the case of a characteristic such as anxiety. The practical significance of this in evaluating a test is that certain factors outside the test itself can serve to reduce the reliability that the test can realistically be expected to achieve. Thus, an examiner should generally expect higher reliabilities for an intelligence test than for a test measuring a personality variable such as anxiety. It is the examiner's responsibility to know what is being measured, especially the degree of variability to be expected in the measured trait.

The second important issue relating to reliability is that psychological testing methods are necessarily imprecise. For the hard sciences, researchers can make direct measurements such as the concentration of a chemical solution, the relative weight of one organism compared with another, or the strength of radiation. In contrast, many constructs in psychology are often measured indirectly. For example, intelligence cannot be perceived directly; it must be inferred by measuring behavior that has been defined as being intelligent. Variability relating to these inferences is likely to produce a certain degree of error resulting from the lack of precision in defining and observing inner psychological constructs. Variability in measurement also occurs simply because people have true (not because of test error) fluctuations in performance between one testing session and the next. Whereas it is impossible to control for the natural variability in human performance, adequate test construction can attempt to reduce the imprecision that is a function of the test itself. Natural human variability and test imprecision make the task of measurement extremely difficult. Although some error in testing is inevitable, the goal of test construction is to keep testing errors within reasonably accepted limits. A high correlation is generally .80 or more, but the variable being measured also changes the expected strength of the correlation. Likewise, the method of determining reliability alters the relative strength of the correlation. Ideally, clinicians should hope for correlations of .90 or higher in tests that are used to make decisions about individuals, whereas a correlation of .70 or more is generally adequate for research purposes.

The purpose of reliability is to estimate the degree of test variance caused by error. The four primary methods of obtaining reliability involve determining (a) the extent to which the test produces consistent results on retesting (test-retest), (b) the relative accuracy of a test at a given time (alternate forms), (c) the internal consistency of the items (split half), and (d) the degree of agreement between two examiners (interscorer). Another way to summarize this is that reliability can be time to time (test-retest), form to form (alternate forms), item to item (split half), or scorer to scorer (interscorer). Although these are the main types of reliability, there is a fifth type, the Kuder-Richardson; like the split half, it is a measurement of the internal consistency of the test items. However, because this method is considered appropriate only for tests that are relatively pure measures of a single variable, it is not covered in this book.

Test-Retest Reliability

Test-retest reliability is determined by administering the test and then repeating it on a second occasion. The reliability coefficient is calculated by correlating the scores obtained by the same person on the two different administrations. The degree of correlation between the two scores indicates the extent to which the test scores can be generalized from one situation to the next. If the correlations are high, the results are less

likely to be caused by random fluctuations in the condition of the examinee or the testing environment. Thus, when the test is being used in actual practice, the examiner can be relatively confident that differences in scores are the result of an actual change in the trait being measured rather than random fluctuation.

A number of factors must be considered in assessing the appropriateness of test-retest reliability. One is that the interval between administrations can affect reliability. Thus, a test manual should specify the interval as well as any significant life changes that the examinees may have experienced such as counseling, career changes, or psychotherapy. For example, tests of preschool intelligence often give reasonably high correlations if the second administration is within several months of the first one. However, correlations with later childhood or adult IQ are generally low because of innumerable intervening life changes. One of the major difficulties with test-retest reliability is the effect that practice and memory may have on performance, which can produce improvement between one administration and the next. This is a particular problem for speeded and memory tests such as those found on the Digit Symbol and Arithmetic subtests of the WAIS-III. Additional sources of variation may be the result of random, short-term fluctuations in the examinee, or of variations in the testing conditions. In general, test-retest reliability is the preferred method only if the variable being measured is relatively stable. If the variable is highly changeable (e.g., anxiety), this method is usually not adequate.

Alternate Forms

The alternate forms method avoids many of the problems encountered with test-retest reliability. The logic behind alternate forms is that, if the trait is measured several times on the same individual by using parallel forms of the test, the different measurements should produce similar results. The degree of similarity between the scores represents the reliability coefficient of the test. As in the test-retest method, the interval between administrations should always be included in the manual as well as a description of any significant intervening life experiences. If the second administration is given immediately after the first, the resulting reliability is more a measure of the correlation between forms and not across occasions. Correlations determined by tests given with a wide interval, such as two months or more, provide a measure of both the relation between forms and the degree of temporal stability.

The alternate forms method eliminates many carryover effects, such as the recall of previous responses the examinee has made to specific items. However, there is still likely to be some carryover effect in that the examinee can learn to adapt to the overall style of the test even when the specific item content between one test and another is unfamiliar. This is most likely when the test involves some sort of problem-solving strategy in which the same principle in solving one problem can be used to solve the next one. An examinee, for example, may learn to use mnemonic aids to increase his or her performance on an alternate form of the WAIS-III Digit Symbol subtest.

Perhaps the primary difficulty with alternate forms lies in determining whether the two forms are actually equivalent. For example, if one test is more difficult than its alternate form, the difference in scores may represent actual differences in the two tests rather than differences resulting from the unreliability of the measure. Because the test constructor is attempting to measure the reliability of the test itself and not

the differences between the tests, this could confound and lower the reliability coefficient. Alternate forms should be independently constructed tests that use the same specifications, including the same number of items, type of content, format, and manner of administration.

A final difficulty is encountered primarily when there is a delay between one administration and the next. With such a delay, the examinee may perform differently because of short-term fluctuations such as mood, stress level, or the relative quality of the previous night's sleep. Thus, an examinee's abilities may vary somewhat from one examination to another, thereby affecting test results. Despite these problems, alternate forms reliability has the advantage of at least reducing, if not eliminating, many carryover effects of the test-retest method. A further advantage is that the alternate test forms can be useful for other purposes, such as assessing the effects of a treatment program or monitoring a patient's changes over time by administering the different forms on separate occasions.

Split Half Reliability

The split half method is the best technique for determining reliability for a trait with a high degree of fluctuation. Because the test is given only once, the items are split in half, and the two halves are correlated. As there is only one administration, it is not possible for the effects of time to intervene as they might with the test-retest method. Thus, the split half method gives a measure of the internal consistency of the test items rather than the temporal stability of different administrations of the same test. To determine split half reliability, the test is often split on the basis of odd and even items. This method is usually adequate for most tests. Dividing the test into a first half and second half can be effective in some cases, but is often inappropriate because of the cumulative effects of warming up, fatigue, and boredom, all of which can result in different levels of performance on the first half of the test compared with the second.

As is true with the other methods of obtaining reliability, the split half method has limitations. When a test is split in half, there are fewer items on each half, which results in wider variability because the individual responses cannot stabilize as easily around a mean. As a general principle, the longer a test is, the more reliable it is because the larger the number of items, the easier it is for the majority of items to compensate for minor alterations in responding to a few of the other items. As with the alternate forms method, differences in content may exist between one half and another.

Interscorer Reliability

In some tests, scoring is based partially on the judgment of the examiner. Because judgment may vary between one scorer and the next, it may be important to assess the extent to which reliability might be affected. This is especially true for projectives and even for some ability tests where hard scorers may produce results somewhat different from easy scorers. This variance in interscorer reliability may apply for global judgments based on test scores such as brain-damaged versus normal, or for small details of scoring such as whether a person has given a shading versus a texture response on the Rorschach. The basic strategy for determining interscorer reliability is to obtain a series of responses from a single client and to have these responses scored by two different individuals. A variation is to have two different examiners test the same client using the same test and

then to determine how close their scores or ratings of the person are. The two sets of scores can then be correlated to determine a reliability coefficient. Any test that requires even partial subjectivity in scoring should provide information on interscorer reliability.

The best form of reliability is dependent on both the nature of the variable being measured and the purposes for which the test is used. If the trait or ability being measured is highly stable, the test-retest method is preferable, whereas split half is more appropriate for characteristics that are highly subject to fluctuations. When using a test to make predictions, the test-retest method is preferable because it gives an estimate of the dependability of the test from one administration to the next. This is particularly true if, when determining reliability, an increased time interval existed between the two administrations. If, on the other hand, the examiner is concerned with the internal consistency and accuracy of a test for a single, one-time measure, either the split half or the alternate forms would be best.

Another consideration in evaluating the acceptable range of reliability is the format of the test. Longer tests usually have higher reliabilities than shorter ones. Also, the format of the responses affects reliability. For example, a true-false format is likely to have a lower reliability than multiple choice because each true-false item has a 50% possibility of the answer being correct by chance. In contrast, each question in a multiple-choice format having five possible choices has only a 20% possibility of being correct by chance. A final consideration is that tests with various subtests or subscales should report the reliability for the overall test as well as for each of the subtests. In general, the overall test score has a significantly higher reliability than its subtests. In estimating the confidence with which test scores can be interpreted, the examiner should take into account the lower reliabilities of the subtests. For example, a Full Scale IQ on the WAIS-III can be interpreted with more confidence than the specific subscale scores.

Most test manuals include a statistical index of the amount of error that can be expected for test scores, which is referred to as the *standard error of measurement* (SEM). The logic behind the SEM is that test scores consist of both truth and error. Thus, there is always noise or error in the system, and the SEM provides a range to indicate how extensive that error is likely to be. The range depends on the test's reliability so that the higher the reliability, the narrower the range of error. The SEM is a standard deviation score so that, for example, a SEM of 3 on an intelligence test would indicate that an individual's score has a 68% chance of being ± 3 IQ points from the estimated true score. This is because the SEM of 3 represents a band extending from −1 to +1 standard deviations above and below the mean. Likewise, there would be a 95% chance that the individual's score would fall within a range of ± 5 points from the estimated true score. From a theoretical perspective, the SEM is a statistical index of how a person's repeated scores on a specific test would fall around a normal distribution. Thus, it is a statement of the relationship among a person's obtained score, his or her theoretically true score, and the test reliability. Because it is an empirical statement of the probable range of scores, the SEM has more practical usefulness than a knowledge of the test reliability. This band of error is also referred to as a *confidence interval.*

The acceptable range of reliability is difficult to identify and depends partially on the variable being measured. In general, unstable aspects (states) of the person produce lower reliabilities than stable ones (traits). Thus, in evaluating a test, the examiner

should expect higher reliabilities on stable traits or abilities than on changeable states. For example, a person's general fund of vocabulary words is highly stable and therefore produces high reliabilities. In contrast, a person's level of anxiety is often highly changeable. This means examiners should not expect nearly as high reliabilities for anxiety as for an ability measure such as vocabulary. A further consideration, also related to the stability of the trait or ability, is the method of reliability that is used. Alternate forms are considered to give the lowest estimate of the actual reliability of a test, while split half provides the highest estimate. Another important way to estimate the adequacy of reliability is by comparing the reliability derived on other similar tests. The examiner can then develop a sense of the expected levels of reliability, which provides a baseline for comparisons. In the example of anxiety, a clinician may not know what is an acceptable level of reliability. A general estimate can be made by comparing the reliability of the test under consideration with other tests measuring the same or a similar variable. The most important thing to keep in mind is that lower levels of reliability usually suggest that less confidence can be placed in the interpretations and predictions based on the test data. However, clinical practitioners are less likely to be concerned with low statistical reliability if they have some basis for believing the test is a valid measure of the client's state at the time of testing. The main consideration is that the sign or test score does not mean one thing at one time and something different at another.

Validity

The most crucial issue in test construction is validity. Whereas reliability addresses issues of consistency, validity assesses what the test is to be accurate about. A test that is valid for clinical assessment should measure what it is intended to measure and should also produce information useful to clinicians. A psychological test cannot be said to be valid in any abstract or absolute sense, but more practically, it must be valid in a particular context and for a specific group of people (Messick, 1995). Although a test can be reliable without being valid, the opposite is not true; a necessary prerequisite for validity is that the test must have achieved an adequate level of reliability. Thus, a valid test is one that accurately measures the variable it is intended to measure. For example, a test comprising questions about a person's musical preference might erroneously state that it is a test of creativity. The test might be reliable in the sense that if it is given to the same person on different occasions, it produces similar results each time. However, it would not be reliable in that an investigation might indicate it does not correlate with other more valid measurements of creativity.

Establishing the validity of a test can be extremely difficult, primarily because psychological variables are usually abstract concepts such as intelligence, anxiety, and personality. These concepts have no tangible reality, so their existence must be inferred through indirect means. In addition, conceptualization and research on constructs undergo change over time requiring that test validation go through continual refinement (G. Smith & McCarthy, 1995). In constructing a test, a test designer must follow two necessary, initial steps. First, the construct must be theoretically evaluated and described; second, specific operations (test questions) must be developed to measure it (S. Haynes et al., 1995). Even when the designer has followed these steps closely and conscientiously, it is sometimes difficult to determine what the test really measures. For

example, IQ tests are good predictors of academic success, but many researchers question whether they adequately measure the concept of intelligence as it is theoretically described. Another hypothetical test that, based on its item content, might seem to measure what is described as musical aptitude may in reality be highly correlated with verbal abilities. Thus, it may be more a measure of verbal abilities than of musical aptitude.

Any estimate of validity is concerned with relationships between the test and some external independently observed event. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council for Measurement in Education [NCME], 1999; G. Morgan, Gliner, & Harmon, 2001) list the three main methods of establishing validity as content-related, criterion-related, and construct-related.

Content Validity

During the initial construction phase of any test, the developers must first be concerned with its content validity. This refers to the representativeness and relevance of the assessment instrument to the construct being measured. During the initial item selection, the constructors must carefully consider the skills or knowledge area of the variable they would like to measure. The items are then generated based on this conceptualization of the variable. At some point, it might be decided that the item content overrepresents, underrepresents, or excludes specific areas, and alterations in the items might be made accordingly. If experts on subject matter are used to determine the items, the number of these experts and their qualifications should be included in the test manual. The instructions they received and the extent of agreement between judges should also be provided. A good test covers not only the subject matter being measured, but also additional variables. For example, factual knowledge may be one criterion, but the application of that knowledge and the ability to analyze data are also important. Thus, a test with high content validity must cover all major aspects of the content area and must do so in the correct proportion.

A concept somewhat related to content validity is face validity. These terms are not synonymous, however, because *content validity* pertains to judgments made by experts, whereas *face validity* concerns judgments made by the test users. The central issue in face validity is test rapport. Thus, a group of potential mechanics who are being tested for basic skills in arithmetic should have word problems that relate to machines rather than to business transactions. Face validity, then, is present if the test looks good to the persons taking it, to policymakers who decide to include it in their programs, and to other untrained personnel. Despite the potential importance of face validity in regard to test-taking attitudes, disappointingly few formal studies on face validity are performed and/or reported in test manuals.

In the past, content validity has been conceptualized and operationalized as being based on the subjective judgment of the test developers. As a result, it has been regarded as the least preferred form of test validation, albeit necessary in the initial stages of test development. In addition, its usefulness has been primarily focused at achievement tests (how well has this student learned the content of the course?) and personnel selection (does this applicant know the information relevant to the potential job?). More recently, it has become used more extensively in personality and clinical assessment (Butcher, Graham, Williams, & Ben-Porath, 1990; Millon, 1994). This has

paralleled more rigorous and empirically based approaches to content validity along with a closer integration to criterion and construct validation.

Criterion Validity

A second major approach to determining validity is criterion validity, which has also been called *empirical* or *predictive* validity. Criterion validity is determined by comparing test scores with some sort of performance on an outside measure. The outside measure should have a theoretical relation to the variable that the test is supposed to measure. For example, an intelligence test might be correlated with grade point average; an aptitude test, with independent job ratings or general maladjustment scores, with other tests measuring similar dimensions. The relation between the two measurements is usually expressed as a correlation coefficient.

Criterion-related validity is most frequently divided into either concurrent or predictive validity. *Concurrent validity* refers to measurements taken at the same, or approximately the same, time as the test. For example, an intelligence test might be administered at the same time as assessments of a group's level of academic achievement. *Predictive validity* refers to outside measurements that were taken some time after the test scores were derived. Thus, predictive validity might be evaluated by correlating the intelligence test scores with measures of academic achievement a year after the initial testing. Concurrent validation is often used as a substitute for predictive validation because it is simpler, less expensive, and not as time consuming. However, the main consideration in deciding whether concurrent or predictive validation is preferable depends on the test's purpose. Predictive validity is most appropriate for tests used for selection and classification of personnel. This may include hiring job applicants, placing military personnel in specific occupational training programs, screening out individuals who are likely to develop emotional disorders, or identifying which category of psychiatric populations would be most likely to benefit from specific treatment approaches. These situations all require that the measurement device provide a prediction of some future outcome. In contrast, concurrent validation is preferable if an assessment of the client's current status is required, rather than a prediction of what might occur to the client at some future time. The distinction can be summarized by asking "Is Mr. Jones maladjusted?" (concurrent validity) rather than "Is Mr. Jones likely to become maladjusted at some future time?" (predictive validity).

An important consideration is the degree to which a specific test can be applied to a unique work-related environment (see Hogan, Hogan, & Roberts, 1996). This relates more to the social value and consequences of the assessment than the formal validity as reported in the test manual (Messick, 1995). In other words, can the test under consideration provide accurate assessments and predictions for the environment in which the examinee is working? To answer this question adequately, the examiner must refer to the manual and assess the similarity between the criteria used to establish the test's validity and the situation to which he or she would like to apply the test. For example, can an aptitude test that has adequate criterion validity in the prediction of high school grade point average also be used to predict academic achievement for a population of college students? If the examiner has questions regarding the relative applicability of the test, he or she may need to undertake a series of specific tasks. The first is to identify

the required skills for adequate performance in the situation involved. For example, the criteria for a successful teacher may include such attributes as verbal fluency, flexibility, and good public speaking skills. The examiner then must determine the degree to which each skill contributes to the quality of a teacher's performance. Next, the examiner has to assess the extent to which the test under consideration measures each of these skills. The final step is to evaluate the extent to which the attribute that the test measures is relevant to the skills the examiner needs to predict. Based on these evaluations, the examiner can estimate the confidence that he or she places in the predictions developed from the test. This approach is sometimes referred to as *synthetic validity* because examiners must integrate or synthesize the criteria reported in the test manual with the variables they encounter in their clinical or organizational settings.

The strength of criterion validity depends in part on the type of variable being measured. Usually, intellectual or aptitude tests give relatively higher validity coefficients than personality tests because there are generally a greater number of variables influencing personality than intelligence. As the number of variables that influences the trait being measured increases, it becomes progressively more difficult to account for them. When a large number of variables are not accounted for, the trait can be affected in unpredictable ways. This can create a much wider degree of fluctuation in the test scores, thereby lowering the validity coefficient. Thus, when evaluating a personality test, the examiner should not expect as high a validity coefficient as for intellectual or aptitude tests. A helpful guide is to look at the validities found in similar tests and compare them with the test being considered. For example, if an examiner wants to estimate the range of validity to be expected for the extraversion scale on the Myers Briggs Type Indicator, he or she might compare it with the validities for similar scales found in the California Personality Inventory and Eysenck Personality Questionnaire. The relative level of validity, then, depends both on the quality of the construction of the test and on the variable being studied.

An important consideration is the extent to which the test accounts for the trait being measured or the behavior being predicted. For example, the typical correlation between intelligence tests and academic performance is about .50 (Neisser et al., 1996). Because no one would say that grade point average is entirely the result of intelligence, the relative extent to which intelligence determines grade point average has to be estimated. This can be calculated by squaring the correlation coefficient and changing it into a percentage. Thus, if the correlation of .50 is squared, it comes out to 25%, indicating that 25% of academic achievement can be accounted for by IQ as measured by the intelligence test. The remaining 75% may include factors such as motivation, quality of instruction, and past educational experience. The problem facing the examiner is to determine whether 25% of the variance is sufficiently useful for the intended purposes of the test. This ultimately depends on the personal judgment of the examiner.

The main problem confronting criterion validity is finding an agreed-on, definable, acceptable, and feasible outside criterion. Whereas for an intelligence test the grade point average might be an acceptable criterion, it is far more difficult to identify adequate criteria for most personality tests. Even with so-called intelligence tests, many researchers argue that it is more appropriate to consider them tests of scholastic aptitude rather than of intelligence. Yet another difficulty with criterion validity is the possibility that the criterion measure will be inadvertently biased. This is referred to

as *criterion contamination* and occurs when knowledge of the test results influences an individual's later performance. For example, a supervisor in an organization who receives such information about subordinates may act differently toward a worker placed in a certain category after being tested. This situation may set up negative or positive expectations for the worker, which could influence his or her level of performance. The result is likely to artificially alter the level of the validity coefficients. To work around these difficulties, especially in regard to personality tests, a third major method must be used to determine validity.

Construct Validity

The method of construct validity was developed in part to correct the inadequacies and difficulties encountered with content and criterion approaches. Early forms of content validity relied too much on subjective judgment, while criterion validity was too restrictive in working with the domains or structure of the constructs being measured. Criterion validity had the further difficulty in that there was often a lack of agreement in deciding on adequate outside criteria. The basic approach of construct validity is to assess the extent to which the test measures a theoretical construct or trait. This assessment involves three general steps. Initially, the test constructor must make a careful analysis of the trait. This is followed by a consideration of the ways in which the trait should relate to other variables. Finally, the test designer needs to test whether these hypothesized relationships actually exist (Foster & Cone, 1995). For example, a test measuring dominance should have a high correlation with the individual accepting leadership roles and a low or negative correlation with measures of submissiveness. Likewise, a test measuring anxiety should have a high positive correlation with individuals who are measured during an anxiety-provoking situation, such as an experiment involving some sort of physical pain. As these hypothesized relationships are verified by research studies, the degree of confidence that can be placed in a test increases.

There is no single, best approach for determining construct validity; rather, a variety of different possibilities exist. For example, if some abilities are expected to increase with age, correlations can be made between a population's test scores and age. This may be appropriate for variables such as intelligence or motor coordination, but it would not be applicable for most personality measurements. Even in the measurement of intelligence or motor coordination, this approach may not be appropriate beyond the age of maturity. Another method for determining construct validity is to measure the effects of experimental or treatment interventions. Thus, a posttest measurement may be taken following a period of instruction to see if the intervention affected the test scores in relation to a previous pretest measure. For example, after an examinee completes a course in arithmetic, it would be predicted that scores on a test of arithmetical ability would increase. Often, correlations can be made with other tests that supposedly measure a similar variable. However, a new test that correlates too highly with existing tests may represent needless duplication unless it incorporates some additional advantage such as a shortened format, ease of administration, or superior predictive validity. Factor analysis is of particular relevance to construct validation because it can be used to identify and assess the relative strength of different psychological traits. Factor analysis can also be used in the design of a test to identify the primary factor or factors measured by a series of different tests. Thus, it can be used to simplify one or

more tests by reducing the number of categories to a few common factors or traits. The factorial validity of a test is the relative weight or loading that a factor has on the test. For example, if a factor analysis of a measure of psychopathology determined that the test was composed of two clear factors that seemed to be measuring anxiety and depression, the test could be considered to have factorial validity. This would be especially true if the two factors seemed to be accounting for a clear and large portion of what the test was measuring.

Another method used in construct validity is to estimate the degree of internal consistency by correlating specific subtests with the test's total score. For example, if a subtest on an intelligence test does not correlate adequately with the overall or Full Scale IQ, it should be either eliminated or altered in a way that increases the correlation. A final method for obtaining construct validity is for a test to converge or correlate highly with variables that are theoretically similar to it. The test should not only show this convergent validity but also have discriminate validity, in which it would demonstrate low or negative correlations with variables that are dissimilar to it. Thus, scores on reading comprehension should show high positive correlations with performance in a literature class and low correlations with performance in a class involving mathematical computation.

Related to discriminant and convergent validity is the degree of sensitivity and specificity an assessment device demonstrates in identifying different categories. *Sensitivity* refers to the percentage of true positives that the instrument has identified, whereas *specificity* is the relative percentage of true negatives. A structured clinical interview might be quite sensitive in that it would accurately identify 90% of schizophrenics in an admitting ward of a hospital. However, it may not be sufficiently specific in that 30% of schizophrenics would be incorrectly classified as either normal or having some other diagnosis. The difficulty in determining sensitivity and specificity lies in developing agreed-on, objectively accurate outside criteria for categories such as psychiatric diagnosis, intelligence, or personality traits.

As indicated by the variety of approaches discussed, no single, quick, efficient method exists for determining construct validity. It is similar to testing a series of hypotheses in which the results of the studies determine the meanings that can be attached to later test scores (Foster & Cone, 1995; Messick, 1995). Almost any data can be used, including material from the content and criterion approaches. The greater the amount of supporting data, the greater is the level of confidence with which the test can be used. In many ways, construct validity represents the strongest and most sophisticated approach to test construction. In many ways, all types of validity can be considered as subcategories of construct validity. It involves theoretical knowledge of the trait or ability being measured, knowledge of other related variables, hypothesis testing, and statements regarding the relationship of the test variable to a network of other variables that have been investigated. Thus, construct validation is a never-ending process in which new relationships always can be verified and investigated.

VALIDITY IN CLINICAL PRACTICE

Although a test may have been found to have a high level of validity during its construction, it does not necessarily follow that the test is also valid in a specific situation

with a particular client. A test can never be valid in any absolute sense because, in practice, numerous variables might affect the test results. A serious issue, then, is the degree of validity generalization that is made. In part, this generalization depends on the similarity between the population used during various stages of test construction and the population and situation that it is being used for in practice. Validity in clinical practice also depends on the extent to which tests can work together to improve each other's accuracy. Some tests thus show incremental validity in that they improve accuracy in increments as increasing numbers of data sources are used. *Incremental validity,* then, refers to the ability of tests to produce information above what is already known. Another important consideration is the ability of the clinician to generate hypotheses, test these hypotheses, and blend the data derived from hypothesis testing into a coherent, integrated picture of the person. Maloney and Ward (1976) refer to this latter approach to validity as *conceptual validity* because it involves creating a conceptually coherent description of the person.

Incremental Validity

For a test to be considered useful and efficient, it must be able to produce accurate results above and beyond the results that could be obtained with greater ease and less expense. If equally accurate clinical descriptions could be obtained through such basic information as biographical data and knowing the referral question, there would be no need for psychological tests. Incremental validity also needs to be evaluated in relation to cost effectiveness. A psychological test might indeed demonstrate incremental validity by increasing the relative proportions of accurate diagnoses, or hit rates, by 2%. However, practitioners need to question whether this small increase in accuracy is worth the extra time involved in administering and interpreting the test. Clinicians might direct their time more productively toward direct treatment.

In the 1950s, one of the theoretical defenses for tests having low reliabilities and validities was that, when used in combination, their accuracy could be improved. In other words, results from a series of different tests could provide checks and balances to correct for inaccurate interpretations. A typical strategy used to empirically test for this was to first obtain biographical data, make interpretations and decisions based on this data, and then test its accuracy based on some outside criterion. Next, a test such as the MMPI could be given; then, the interpretations and decisions based on it could likewise be assessed for accuracy. Finally, clinicians could be given both sets of data to assess any improvements in the accuracies of interpretation/decisions between either of the first two conditions and the combined information.

It would seem logical that the greater the number of tests used, the greater would be the overall validity of the assessment battery. However, research on psychological tests used in clinical practice has often demonstrated that they have poor incremental validity. An older but representative study by Kostlan (1954) on male psychiatric outpatients compared the utility of a case history, Rorschach, MMPI, and a sentence completion test. Twenty experienced clinicians interpreted different combinations of these sources of test data. Their conclusions were combined against criterion judges who used a lengthy checklist of personality descriptions. The conclusions were that, for most of the data, the clinicians were no more accurate than if they had used only age, occupation, education, marital status, and a basic description of the referral

question. The exception was that the most accurate descriptions were based on a combination of social history and the MMPI. In contrast, psychological tests have sometimes clearly demonstrated their incremental validity. S. Schwartz and Wiedel (1981) demonstrated that neurological residents gave more accurate diagnoses when an MMPI was used in combination with history, electroencephalogram (EEG), and physical exam. This was probably not so much because of a specific MMPI neurological profile, but rather that the MMPI increased diagnostic accuracy by enabling the residents to rule out other possible diagnoses.

Often clinical psychologists attempt to make a series of behavioral predictions based on complex psychological tests. Although these predictions may show varying levels of accuracy, a simpler and more effective means of achieving this information might be to simply ask the clients to predict their own behaviors. In some circumstances, self-prediction has been found to be more accurate than psychological tests, whereas in others, tests have been found to be more accurate (Shrauger & Osberg, 1981). Advantages of self-assessment are that it can be time-efficient, cost-effective, and facilitate a collegial relationship between assessor and client. In contrast, difficulties are that, compared with formal testing, self-assessment may be significantly more susceptible to social desirability, attributional errors, distortions caused by poor adjustment, and the relative self-awareness of the client. These factors need to be carefully considered before deciding to use self-assessment versus formal psychological tests. Although the incremental validity of using self-assessment in combination with formal testing has not been adequately researched, it would seem that this is conceptually a potentially useful strategy for future research.

Reviews of studies on incremental validity (Garb, 1998b) have provided a number of general conclusions. The addition of an MMPI to background data has consistently led to increases in validity although the increases were quite small when the MMPI was added to extensive data. The addition of projective tests to a test battery did not generally increase incremental validity. Lanyon and Goodstein (1982) have argued that case histories are generally preferable to psychological test data. Furthermore, a single test in combination with case history data is generally as effective as a large number of tests with case history data. Some studies have found that the MMPI alone was generally found to be preferable to a battery containing the MMPI, Rorschach, and sentence completion (Garb, 1984, 1994a, 1998b). In contrast, other studies have found that the Rorschach can add incremental validity to a test battery (G. Meyer, 1997; Weiner, 1999).

In defense of the poor incremental validity of many of the traditional clinical tests are weaknesses and unanswered questions relating to the preceding research. First, few studies have looked at statistically derived predictions and interpretations based on optimal multiple cutoff scores or multiple regression equations. However, more recent research, particularly on tests such as the MMPI and CPI, has emphasized this approach. For example, combined weightings on such variables as specific CPI scores, Scholastic Aptitude Test (SAT) scores, grade point average (GPA), and IQ can be combined to predict success in specific programs (see Chapter 9). Further research using this approach may yield greater incremental validity for a wide number of assessment techniques. Second, few studies on incremental validity have investigated the ways in which different tests might show greater incremental validity in specific situations for specific populations. Instead, most research has focused on the validity of global personality descriptions, without tying these descriptions to the unique circumstances or

contexts persons might be involved in. Finally, as most previous studies have focused on global personality descriptions, certain tests demonstrate greater incremental validity when predicting highly specific traits and behaviors.

Conceptual Validity

A further method for determining validity that is highly relevant to clinical practice is conceptual validity (Maloney & Ward, 1976). In contrast to the traditional methods (content validity, etc.), which are primarily concerned with evaluating the theoretical constructs in the test itself, conceptual validity focuses on individuals with their unique histories and behaviors. It is a means of evaluating and integrating test data so that the clinician's conclusions make accurate statements about the examinee. There are similarities with construct validity in that construct validity also tries to test specific hypothesized relationships between constructs. Conceptual validity is likewise concerned with testing constructs, but in this case the constructs relate to the individual rather than to the test itself.

In determining conceptual validity, the examiner generally begins with individuals for whom no constructs have been developed. The next phase is to observe, collect data, and form a large number of hypotheses. If these hypotheses are confirmed through consistent trends in the test data, behavioral observations, history, and additional data sources, the hypotheses can be considered to represent valid constructs regarding the person. The focus is on an individual in his or her specific situation, and the data are derived from a variety of sources. The conceptual validity of the constructs is based on the logicalness and internal consistency of the data. Unlike construct validity, which begins with previously developed constructs, conceptual validity produces constructs as its end product. Its aim is for these constructs to provide valid sources of information that can be used to help solve the unique problems that an individual may be facing.

CLINICAL JUDGMENT

Any human interaction involves mutual and continually changing perceptions. *Clinical judgment* is a special instance of perception in which the clinician attempts to use whatever sources are available to create accurate descriptions of the client. These sources may include test data, case history, medical records, personal journals, and verbal and nonverbal observations of behavior. Relevant issues and processes involved in clinical judgment include data gathering, data synthesis, the relative accuracy of clinical versus statistical/actuarial descriptions, and judgment in determining what to include in a psychological report. This sequence also parallels the process clinicians go through when assessing a client.

Data Gathering and Synthesis

Most of the research related to the strengths and weaknesses of data gathering and synthesis has focused on the assessment interview (see Chapter 3). However, many of the issues and problems related to clinical judgment during interviewing also have implications

for the gathering and synthesis of test data. One of the most essential elements in gathering data from any source is the development of an optimum level of rapport. Rapport increases the likelihood that clients will give their optimum level of performance. If rapport is not sufficiently developed, it is increasingly likely that the data obtained from the person will be inaccurate.

Another important issue is that the interview itself is typically guided by the client's responses and the clinician's reaction to these responses. A client's responses might be nonrepresentative because of factors such as a transient condition (stressful day, poor night's sleep, etc.) or conscious/unconscious faking. The client's responses also need to be interpreted by the clinician. These interpretations can be influenced by a combination of personality theory, research data, and the clinician's professional and personal experience. The clinician typically develops hypotheses based on a client's responses and combines his or her observations with his or her theoretical understanding of the issue. These hypotheses can be further investigated and tested by interview questions and test data, which can result in confirmation, alteration, or elimination of the hypotheses. Thus, bias can potentially enter into this process from a number of different directions, including the types of questions asked, initial impressions, level of rapport, or theoretical perspective.

The clinician typically collects much of the initial data regarding a client through unstructured or semistructured interviews. Unstructured approaches in gathering and interpreting data provide flexibility, focus on the uniqueness of the person, and are ideographically rich. In contrast, an important disadvantage of unstructured approaches is that a clinician, like most other persons, can be influenced by a number of personal and cultural biases. For example, clinicians might develop incorrect hypotheses based on first impressions (primacy effect). They might end up seeking erroneous confirmation of incorrect hypotheses by soliciting expected responses rather than objectively probing for possible disconfirmation. Thus, clinicians might be unduly influenced by their preferred theory of personality, halo effects, self-fulfilling prophecies, expectations, and cultural stereotypes. These areas of potential sources of error have led to numerous questions regarding the dependability of clinical judgment.

Accuracy of Clinical Judgments

After collecting and organizing their data, clinicians then need to make final judgments regarding the client. Determining the relative accuracy of these judgments is crucial. In some cases, clinical judgment is clearly in error, whereas in others it can be quite accurate. To increase accuracy, clinicians need to know how errors might occur, how to correct these errors, and the relative advantages of specialized training.

A possible source of inaccuracy is that clinicians frequently do not take into account the base rate, or the rate at which a particular behavior, trait, or diagnosis occurs in the general population (Faust, 1991; S. Hawkins & Hastie, 1990; Wedding & Faust, 1989). For example, an intake section of a psychiatric hospital might evaluate a population of whom 50% could be considered to be schizophrenic. A clinician who would randomly diagnose patients as either schizophrenic or nonschizophrenic would be correct 50% of the time. Thus, even a 60% correct diagnosis of schizophrenia would exceed the base rate (or chance occurrence) by only 10%. It is also rare for clinicians to receive feedback

regarding either the accuracy of their diagnoses or other frequently used judgments such as behavioral predictions, personality traits, or the relative success of their recommendations (Garb, 1989, 1994a, 1998b). Thus, it is possible that inaccurate strategies for arriving at conclusions will be continued with little likelihood of correction.

A further source of error is that information obtained earlier in the data collection process is frequently given more importance than information received later (primacy effect). This means that different starting points in the decision-making process may result in different conclusions. This can be further reinforced if clinicians make early judgments and then work to confirm these judgments through seeking supporting information. This confirmatory bias can be especially likely to occur in a hypothesis-testing situation in which clinicians do not adequately seek information that could disconfirm as well as confirm their hypothesis (Haverkamp, 1993). The most problematic examples occur when clinicians interpret a client's behavior and then work to persuade the client that their interpretation is correct (Loftus, 1993).

Research on person perception accuracy indicates that, even though no two persons are uniformly accurate, some persons are much better at accurately perceiving others. Taft (1955) and P. E. Vernon (1964) summarize the early research on person perception accuracy by pointing out that accuracy is not associated with age (in adults); there is little difference in accuracy between males and females (although females are slightly better); and accurate perceptions of others are positively associated with intelligence, artistic/dramatic interests, social detachment, and good emotional adjustment. Authoritarian personalities tend to be poor judges. In most instances, accuracy is related to similarity in race and cultural backgrounds (P. Shapiro & Penrod, 1986). In some cases, accuracy by psychologists may be only slightly related to their amount of clinical experience (Garb, 1989, 1992, 1994a, 1998b); and, for some judgments, psychologists may be no better than certain groups of nonprofessionals, such as physical scientists and personnel workers (Garb, 1992, 1994a, 1998b). Relatively higher rates of accuracy were achieved when clinical judgments based on interviews were combined with formal assessments and when statistical interpretive rules were used. When subjective test interpretation was combined with clinical judgment, it was questionable whether any increase in accuracy was obtained (Garb, 1984, 1989).

It would be logical to assume that the more confidence clinicians feel regarding the accuracy of their judgments, the more likely it would be that their judgments would be accurate. In several studies, however, confidence was often not related to accuracy (E. Kelly & Fiske, 1951; Kleinmuntz, 1990). Kelly and Fiske even found that degree of confidence was inversely related to predicting the success of trainees in a Veterans Administration training program. Several studies (Kareken & Williams, 1994; Lichtenstein & Fischoff, 1977) concluded that persons were generally overconfident regarding judgments; and when outcome knowledge was made available, clinicians typically overestimated what they thought they knew before receiving outcome knowledge (Hawkins & Hastie, 1990). This is usually referred to as *hindsight bias* ("I would have known it all along") and is usually accompanied by a denial that the outcome knowledge has influenced judgment. Paradoxically, as knowledge and experience in an area increase, there is generally a decrease in confidence regarding judgments. This observation was found to be true unless the clinicians were very knowledgeable, in which case they were likely to have a moderate level of confidence (Garb, 1989). Confidence

was also more accurate if participants were made socially accountable for their judgments (Ruscio, 2000). Thus, the more experienced clinicians and persons who were more socially accountable were able to more accurately rate their level of confidence.

Crucial to clinical judgment is whether clinicians can make judgments better than laypersons and whether amount of clinical training can increase accuracy. This is a particularly important issue if psychologists are offering their services as expert witnesses to the legal justice system. Research reviews generally support the value of clinical training, but this is dependent on the domain being assessed. For example, Garb (1992) has concluded, "Clinicians are able to make reliable and valid judgments for many tasks, and their judgments are frequently more valid than judgments by laypersons" (p. 451). In particular, clinicians have been found to make more accurate judgments relating to relatively complex technical areas such as clinical diagnosis, ratings of mental status, many domains related to interview information, short-term (and possibly long-term) predictions of violence, psychological test interpretation (WAIS, MMPI), forensic knowledge, competency evaluations, neuropsychological test results, psychotherapy data, and biographical data (see primarily Garb, 1998b, but also 1984, 1989, 1992, 1994a). In contrast, trained clinicians were no better than laypersons in making judgments based on projective test results and in making personality descriptions based on face-to-face interaction.

The preceding material indicates that errors in clinical judgment can and do occur. It is thus crucial, especially when appearing as an expert witness, that clinicians be familiar with the relevant literature on clinical judgment and, based on this information, take steps to improve their accuracy. Accordingly, Garb (1994a, 1998b) and Wedding and Faust (1989) have made the following recommendations:

1. To avoid missing crucial information, clinicians should use comprehensive, structured, or at least semistructured approaches to interviewing. This is especially important in cases where urgent clinical decisions (danger to self or others) may need to occur.
2. Clinicians should not only consider the data that supports their hypotheses, but also carefully consider or even list evidence that does not support their hypotheses. This will likely reduce the possibility of hindsight and confirmatory bias.
3. Diagnoses should be based on careful attention to the specific criteria contained in the *DSM-IV-TR* (2000; or *International Classification of Disorders [ICD-10]*). In particular, this means not making errors caused by inferences biased by gender and ethnicity.
4. Because memory can be a reconstructive process subject to possible errors, clinicians should avoid relying on memory and, rather, refer to careful notes as much as possible.
5. In making predictions, clinicians should attend to base rates as much as possible. Such a consideration potentially provides a rough estimate of how frequently the behavior will occur in a given population or context. Any clinical predictions, then, are guided by this base rate occurrence and are likely to be improvements on the base rate.
6. Clinicians should seek feedback when possible regarding the accuracy and usefulness of their judgments. For example, psychological reports should ideally be followed up with rating forms (that can be completed by the referral sources) relating to

the clarity, precision, accuracy, and usefulness of the information and recommendations contained in the reports (see Ownby & Wallbrown, 1983).

7. Clinicians should learn as much as possible regarding the theoretical and empirical material relevant to the person or group they are assessing. This would potentially help to develop strategies for obtaining comprehensive information, allow clinicians to make correct estimates regarding the accuracy of their judgments, and provide them with appropriate base rate information.

8. Familiarity with the literature on clinical judgment should be used to continually update practitioners on past and emerging trends.

Sometimes in court proceedings, psychologists are challenged regarding the difficulties associated with clinical judgment. If the preceding steps are taken, psychologists can justifiably reply that they are familiar with the literature and have taken appropriate steps to guard against inaccuracies in clinical judgment. More importantly, the quality of service related to clients and referral sources is also likely to be enhanced.

Clinical versus Actuarial Prediction

Nearly 50 years ago, Meehl (1954) published a review of research comparing the relative accuracy of clinical judgment to statistical formulas when used on identical sets of data (life history, demographic data, test profiles). The clinical approach used clinicians' judgment, whereas the actuarial approach used empirically derived formulas, such as single/multiple cutoffs and regression equations, to come to decisions regarding a client. His review covered a large number of settings including military placement, college success, criminal recidivism, and benefit from psychotherapy. He concluded that statistical decisions consistently outperformed clinical judgments (Meehl, 1954, 1965). This resulted in some lively debate in the journals, with Meehl's conclusions generally being supported (Garb, 1994b; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Kleinmuntz, 1990). Dawes and Corrigan (1974) even found that an actuarial formula based on specific clinicians' own decision-making processes yielded more valid future predictions than the clinicians' own predictions. This was probably because of the formula reducing the influence of uncontrolled errors in the clinicians' procedures.

Despite the empirical support for an actuarial approach, several practical and theoretical issues need to be considered. A clinical approach to integrating data and arriving at conclusions allows a clinician to explore, probe, and deepen his or her understanding in many areas. These frequently involve areas that tests or statistical formulas cannot measure. Often an interview is the only means of obtaining observations of behavior and unique aspects of history. Idiosyncratic events with a low frequency of occurrence may significantly alter a clinician's conclusions although no formulas take these events into account. It is quite common for unique, rare events to have occurred at some time in a client's life; and, during the process of assessment, they are frequently relevant and can often alter the conclusions of many, if not most, clinical assessments. Not only do unique aspects of a person change interpretations, but typically an assessment for a person needs to be focused for a specific context and specific situation that he or she is involved in. When the focus changes from institutional to individual decision making, the

relevance of statistical rules becomes less practical (McGrath, 2001; Vane & Guarnaccia, 1989). Not only are individuals too multifaceted, but also their unique situations, contexts, and the decisions facing them are even more multifaceted.

A further difficulty with a purely actuarial approach is that development of both test reliability and validity, as well as actuarial formulas, requires conceiving the world as stable and static. For such approaches to be useful, the implicit assumption is that neither people nor criteria change. In contrast, the practitioner must deal with a natural world that is imperfect, constantly changing, does not necessarily follow rules, is filled with constantly changing perceptions, and is subject to chance or at least impossible-to-predict events. Thus, even when statistical formulas are available, they may not apply. This distinction between the statistical orientation of the psychometrician and the natural environment of the practitioner underlies the discrepancy between their two worlds (Beutler, 2000). Practitioners must somehow try to combine these two modes of analysis, but often find the task difficult. It may be true that controlled studies generally favor a statistical approach over a clinical one but, at the same time, that truth is seldom useful to the practitioner involved in the changing and unique world of practice (Bonarius, 1984). Often, there is no alternative other than to rely on clinical judgment to combine a wide variety of relevant information. This return to a pre-Meehl perspective is unfortunate and is accepted by most clinicians with hesitation.

Bonarius (1984) presents a conceptual alternative to this dilemma. The first step is to alter mechanistic views of prediction. Instead, clinicians might avoid the term *prediction* altogether and use *anticipation.* Anticipating future possibilities implies a cognitive constructional process rather than a mechanical process. It admits that the world can never be perfect in any mechanistic sense and that there is no such thing as an average person in an average situation engaged in an average interaction. Furthermore, the creation of future events is shared by coparticipants. Clients take an active part in formulating and evaluating their goals. The success of future goals depends on the degree of effort they are willing to put into them. The coparticipants share responsibility for the future. Thus, the likelihood that future events will occur is related to both cognitive constructions of an idiosyncratic world and interaction between participants.

Ideally, clinicians need to be aware of and to use, whenever available, actuarial approaches such as multiple cutoffs and regression equations. Future computer-assisted analysis of assessment results can increasingly provide actuarial predictions especially from multiple sources (i.e., Garb, 2000; Groth-Marnat, 2000b). The conclusions reached from actuarial approaches also need to be integrated with data and inferences obtainable only through clinical means. If unusual details regarding a client are discovered and result in altering an interpretation, the basis for this alteration should be noted in the psychological report. Clinicians should also be sensitive to individual differences in person perception accuracy between one practitioner and the next. These differences may depend on experience, training, knowledge, personality, and the amount and quality of feedback regarding the perceptions of different clinicians. In addition, clinicians must recognize possible increases and decreases in test interpretation and clinical judgment resulting from the incremental validity of their instruments because more information does not necessarily increase the accuracy of clinically based predictions (Garb, 1994b, 1998b; Kleinmuntz, 1990). While it is unlikely that actuarial prediction

rules will replace clinical judgment, formal prediction rules can and should be used more extensively as a resource to improve the accuracy of clinical decision making.

Psychological Report

An accurate and effective psychological report requires that clinicians clarify their thinking and crystallize their interpretations. The report ties together all sources of information, often combining complex interprofessional and interpersonal issues. All the advantages and limitations involved with clinical judgment either directly or indirectly affect the report. The focus should be a clear communication of the clinician's interpretations, conclusions, and recommendations. Chapter 15 provides in-depth information on the psychological report as it relates to relevant research, guidelines, format, and sample reports.

PHASES IN CLINICAL ASSESSMENT

An outline of the phases of clinical assessment can provide both a conceptual framework for approaching an evaluation and a summary of some of the points already discussed. Although the steps in assessment are isolated for conceptual convenience, in actuality, they often occur simultaneously and interact with one another. Throughout these phases, the clinician should integrate data and serve as an expert on human behavior rather than merely an interpreter of test scores. This is consistent with the belief that a psychological assessment can be most useful when it addresses specific individual problems and provides guidelines for decision making regarding these problems.

Evaluating the Referral Question

Many of the practical limitations of psychological evaluations result from an inadequate clarification of the problem. Because clinicians are aware of the assets and limitations of psychological tests, and because clinicians are responsible for providing useful information, it is their duty to clarify the requests they receive. Furthermore, they cannot assume that initial requests for an evaluation are adequately stated. Clinicians may need to uncover hidden agendas, unspoken expectations, and complex interpersonal relationships, as well as explain the specific limitations of psychological tests. One of the most important general requirements is that clinicians understand the vocabulary, conceptual model, dynamics, and expectations of the referral setting in which they will be working (Turner et al., 2001).

Clinicians rarely are asked to give a general or global assessment, but instead are asked to answer specific questions. To address these questions, it is sometimes helpful to contact the referral source at different stages in the assessment process. For example, it is often important in an educational evaluation to observe the student in the classroom environment. The information derived from such an observation might be relayed back to the referral source for further clarification or modification of the referral question. Likewise, an attorney may wish to somewhat alter his or her referral question based on preliminary information derived from the clinician's initial interview with the client.

Acquiring Knowledge Relating to the Content of the Problem

Before beginning the actual testing procedure, examiners should carefully consider the problem, the adequacy of the tests they will use, and the specific applicability of that test to an individual's unique situation. This preparation may require referring both to the test manual and to additional outside sources. Clinicians should be familiar with operational definitions for problems such as anxiety disorders, psychoses, personality disorders, or organic impairment so that they can be alert to their possible expression during the assessment procedure. Competence in merely administering and scoring tests is insufficient to conduct effective assessment. For example, the development of an IQ score does not necessarily indicate that an examiner is aware of differing cultural expressions of intelligence or of the limitations of the assessment device. It is essential that clinicians have in-depth knowledge about the variables they are measuring or their evaluations are likely to be extremely limited.

Related to this is the relative adequacy of the test in measuring the variable being considered. This includes evaluating certain practical considerations, the standardization sample, and reliability and validity (see Table 1.1). It is important that the examiner also consider the problem in relation to the adequacy of the test and decide whether a specific test or tests can be appropriately used on an individual or group. This demands knowledge in such areas as the client's age, sex, ethnicity, race, educational background, motivation for testing, anticipated level of resistance, social environment, and interpersonal relationships. Finally, clinicians need to assess the effectiveness or utility of the test in aiding the treatment process.

Data Collection

After clarifying the referral question and obtaining knowledge relating to the problem, clinicians can then proceed with the actual collection of information. This may come from a wide variety of sources, the most frequent of which are test scores, personal history, behavioral observations, and interview data. Clinicians may also find it useful to obtain school records, previous psychological observations, medical records, police reports, or discuss the client with parents or teachers. It is important to realize that the tests themselves are merely a single tool, or source, for obtaining data. The case history is of equal importance because it provides a context for understanding the client's current problems and, through this understanding, renders the test scores meaningful. In many cases, a client's history is of even more significance in making predictions and in assessing the seriousness of his or her condition than his or her test scores. For example, a high score on depression on the MMPI-2 is not as helpful in assessing suicide risk as are historical factors such as the number of previous attempts, age, sex, details regarding any previous attempts, and length of time the client has been depressed. Of equal importance is that the test scores themselves are usually not sufficient to answer the referral question. For specific problem solving and decision making, clinicians must rely on multiple sources and, using these sources, check to assess the consistency of the observations they make.

Interpreting the Data

The end product of assessment should be a description of the client's present level of functioning, considerations relating to etiology, prognosis, and treatment recommendations. Etiologic descriptions should avoid simplistic formulas and should instead focus on the influence exerted by several interacting factors. These factors can be divided into primary, predisposing, precipitating, and reinforcing causes, and a complete description of etiology should take all of these into account. Further elaborations may also attempt to assess the person from a systems perspective in which the clinician evaluates patterns of interaction, mutual two-way influences, and the specifics of circular information feedback. An additional crucial area is to use the data to develop an effective plan for intervention (see Beutler & Clarkin, 1990; Beutler, Clarkin, & Bongar, 2000; Jongsma & Peterson, 1995). Clinicians should also pay careful attention to research on, and the implications of, incremental validity and continually be aware of the limitations and possible inaccuracies involved in clinical judgment. If actuarial formulas are available, they should be used when possible. These considerations indicate that the description of a client should not be a mere labeling or classification, but should rather provide a deeper and more accurate understanding of the person. This understanding should allow the examiner to perceive new facets of the person in terms of both his or her internal experience and his or her relationships with others.

To develop these descriptions, clinicians must make inferences from their test data. Although such data is objective and empirical, the process of developing hypotheses, obtaining support for these hypotheses, and integrating the conclusions is dependent on the experience and training of the clinician. This process generally follows a sequence of developing impressions, identifying relevant facts, making inferences, and supporting these inferences with relevant and consistent data. Maloney and Ward (1976) have conceptualized a seven-phase approach (Figure 1.1) to evaluating data. They note that, in actual practice, these phases are not as clearly defined as indicated in Figure 1.1, but often occur simultaneously. For example, when a clinician reads a referral question or initially observes a client, he or she is already developing hypotheses about that person and checking to assess the validity of these observations.

Phase 1

The first phase involves collecting data about the client. It begins with the referral question and is followed by a review of the client's previous history and records. At this point, the clinician is already beginning to develop tentative hypotheses and to clarify questions for investigation in more detail. The next step is actual client contact, in which the clinician conducts an interview and administers a variety of psychological tests. The client's behavior during the interview, as well as the content or factual data, is noted. Out of this data, the clinician begins to make his or her inferences.

Phase 2

Phase 2 focuses on the development of a wide variety of inferences about the client. These inferences serve both a summary and explanatory function. For example, an examiner may infer that a client is depressed, which also may explain his or her slow

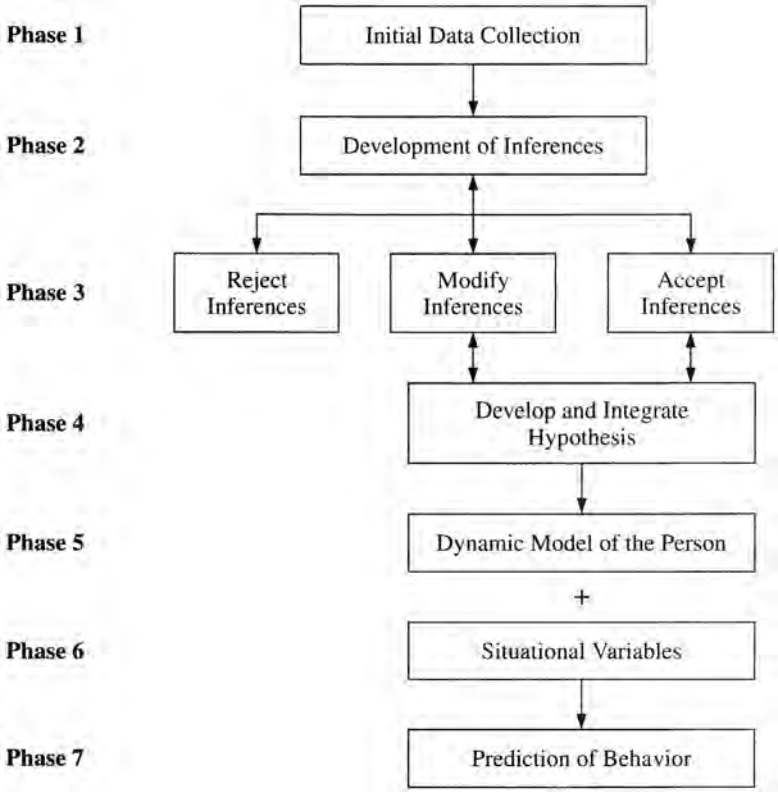


Figure 1.1 Conceptual model for interpreting assessment data

Adapted from Maloney and Ward, 1976, p. 161. Reprinted by permission from *Psychological Assessment: A Conceptual Approach,* by M. P. Maloney and M. P. Ward, New York: Oxford University Press, 1976.

performance, distractibility, flattened affect, and withdrawn behavior. The examiner may then wish to evaluate whether this depression is a deeply ingrained trait or more a reaction to a current situational difficulty. This may be determined by referring to test scores, interview data, or any additional sources of available information. The emphasis in the second phase is on developing multiple inferences that should initially be tentative. They serve the purpose of guiding future investigation to obtain additional information that is then used to confirm, modify, or negate later hypotheses.

Phase 3

Because the third phase is concerned with either accepting or rejecting the inferences developed in Phase 2, there is constant and active interaction between these phases. Often, in investigating the validity of an inference, a clinician alters either the meaning or the emphasis of an inference, or develops entirely new ones. Rarely is an inference entirely substantiated, but rather the validity of that inference is progressively strengthened as the clinician evaluates the degree of consistency and the strength of data that support a particular inference. For example, the inference that a client is

anxious may be supported by WAIS-III subscale performance, MMPI-2 scores, and behavioral observations, or it may only be suggested by one of these sources. The amount of evidence to support an inference directly affects the amount of confidence a clinician can place in this inference.

Phase 4

As a result of inferences developed in the previous three phases, the clinician can move in Phase 4 from specific inferences to general statements about the client. This involves elaborating each inference to describe trends or patterns of the client. For example, the inference that a client is depressed may result from self-verbalizations in which the client continually criticizes and judges his or her behavior. This may also be expanded to give information regarding the ease or frequency with which a person might enter into the depressive state. The central task in Phase 4 is to develop and begin to elaborate on statements relating to the client.

Phases 5, 6, 7

The fifth phase involves a further elaboration of a wide variety of the personality traits of the individual. It represents an integration and correlation of the client's characteristics. This may include describing and discussing general factors such as cognitive functioning, affect and mood, and interpersonal-intrapersonal level of functioning. Although Phases 4 and 5 are similar, Phase 5 provides a more comprehensive and integrated description of the client than Phase 4. Finally, Phase 6 places this comprehensive description of the person into a situational context and Phase 7 makes specific predictions regarding his or her behavior. Phase 7 is the most crucial element involved in decision making and requires that the clinician take into account the interaction between personal and situational variables.

Establishing the validity of these inferences presents a difficult challenge for clinicians because, unlike many medical diagnoses, psychological inferences cannot usually be physically documented. Furthermore, clinicians are rarely confronted with feedback about the validity of these inferences. Despite these difficulties, psychological descriptions should strive to be reliable, have adequate descriptive breadth, and possess both descriptive and predictive validity. Reliability of descriptions refers to whether the description or classification can be replicated by other clinicians (interdiagnostician agreement) as well as by the same clinician on different occasions (intradiagnostician agreement). The next criterion is the breadth of coverage encompassed in the classification. Any classification should be broad enough to encompass a wide range of individuals, yet specific enough to provide useful information regarding the individual being evaluated. Descriptive validity involves the degree to which individuals who are classified are similar on variables external to the classification system. For example, are individuals with similar MMPI-2 profiles also similar on other relevant attributes such as family history, demographic variables, legal difficulties, or alcohol abuse? Finally, predictive validity refers to the confidence with which test inferences can be used to evaluate future outcomes. These may include academic achievement, job performance, or the outcome of treatment. This is one of the most crucial functions of testing. Unless inferences can be made that effectively enhance decision making, the scope and relevance of testing

are significantly reduced. Although these criteria are difficult to achieve and to evaluate, they represent the ideal standard for which assessments should strive.

RECOMMENDED READING

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment.* Washington, DC: American Psychological Association.

Groth-Marnat, G. (2000). Visions of clinical assessment: Then, now, and a brief history of the future. *Journal of Clinical Psychology, 56,* 349–365.

Kaplan, R. M., & Saccuzzo, D. (2001). *Psychological testing: Principles, applications, and issues* (5th ed.). Pacific Grove, CA: Wadsworth.

Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., et al. (2000). Empirical support for psychological assessment in clinical care settings. *Professional Psychology, 31,* 119–130.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist, 45,* 999–1017.

Meyer, G. J., Finn, S. E., Eyde, L., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment. *American Psychologist, 56,* 128–165.

Chapter 2

CONTEXT OF CLINICAL ASSESSMENT

Although general knowledge regarding tests and test construction is essential, practitioners must consider a wide range of additional issues to place testing procedures and test scores in an appropriate context. These considerations include clarifying the referral question, understanding the referral context, following ethical guidelines, identifying and working with test bias, selecting the most appropriate instrument for the variable or problem being studied, and making appropriate use of computer-assisted interpretation.

TYPES OF REFERRAL SETTINGS

Throughout the assessment process, practitioners should try to understand the unique problems and demands encountered in different referral settings. Otherwise, examiners—despite being skilled in administering and interpreting tests—may provide much useless information to their referral source and perhaps even administer a needless series of tests. That is, a thorough investigation of the underlying motive for a referral can sometimes lead to the discovery that evaluation through testing is not warranted.

Errors in test interpretation frequently occur because clinicians do not respond to the referral question in its broadest context. In turn, requests for psychological testing are often worded vaguely: "I would like a psychological evaluation on Mr. Smith," or "Could you evaluate Jimmy because he is having difficulties in school?" The request seldom states a specific question that must be answered or a decision that must be made, when in fact this is usually the position that the referral source is in. For example, a school administrator may need testing to support a placement decision, a teacher may want to prove to parents that their child has a serious problem, or a psychiatric resident may not be comfortable with the management of a patient. An organization's surface motive for testing may be as vague as a statement that the procedure is a matter of policy. Greater clarification is necessary before clinicians can provide useful problem-solving information. Furthermore, many of these situations have hidden agendas that may not be adequately handled through psychological testing alone.

It must be stressed that the responsibility for exploring and clarifying the referral question lies with the clinician who should actively work with the referral source to place the client's difficulty in a practicable context. Clinicians must understand the decisions that the referral source is facing, as well as the available alternatives and the relative usefulness of each of these alternatives. Clinicians also need to specify the relevance of the psychological evaluation in determining different alternatives and their

possible outcomes. They should make clear the advantages and usefulness of psychological testing, but should also explain the limitations inherent in test data.

To help clarify the referral question as well as develop a relevant psychological evaluation, clinicians should become familiar with the types of environments in which they will be working. The most frequent environments are the psychiatric setting, the general medical setting, the legal context, the educational context, and the psychological clinic.

Psychiatric Setting

Levine (1981) has summarized the important factors for a psychologist to be aware of in a psychiatric setting. These referrals typically come from a psychiatrist, who may be asking the referral question in the role of administrator, psychotherapist, or physician. Each role presents unique issues for the psychiatrist, and clinicians have a primary responsibility to develop evaluations that directly address the problems at hand.

One of the main roles a psychiatrist fills is administrator in a ward. Ward administrators frequently must make decisions about problems such as suicide risk, admission/discharge, and the suitability of a wide variety of medical procedures. While retaining primary responsibility, a psychiatrist often uses information from other persons to help with decisions. This represents a change from the typical role of psychiatrists 30 years ago when psychiatrists were mainly concerned with diagnosis and treatment. Currently, issues about custody, freedom of the patient, and the safety of society have taken over as the primary focus. From the perspective of psychologists performing assessments, this means that making a formal *DSM-IV* (1994) psychiatric diagnosis is usually not sufficient in and of itself. For example, a patient may be diagnosed manic-depressive, but this label does not indicate the level of dangerousness that the patient poses to himself or herself or to others. After patients have been admitted to a psychiatric setting, many practical questions have to be answered, such as the type of ward in which to place them, the activities in which they should be involved, and the method of therapy that would be most likely to benefit them.

Initially, the psychologist must determine exactly what information the ward administrator is looking for, particularly concerning any decisions that must be made about the patient. Psychologists in psychiatric settings who receive vague requests for "a psychological" sometimes develop a standard evaluation based on their preconception of what this term implies. They may evaluate the patient's defense mechanisms, diagnosis, cognitive style, and psychosocial history without addressing the specific decisions that have to be made or perhaps covering only two or three relevant issues and omitting others. To maximize the usefulness of an evaluation, examiners must be especially aware of, and sensitive to, psychiatric administrators' legal and custodial responsibilities.

In contrast to the concerns of ward administrators, the standard referral questions from psychiatrists evaluating a patient for possible psychotherapy involve the appropriateness of the client for such therapy, the strategies that are most likely to be effective, and the likely outcome of therapy. These assessments are usually clear-cut and typically do not present any difficulties. Such an evaluation can elaborate on likely problems that may occur during the course of therapy, capacity for insight, diagnosis, coping style, level of resistance, degree of functional impairment, and problem complexity (see Chapter 14).

If a referral is made during therapy, however, a number of problem areas may exist that are not readily apparent from the referral question. The assessor must investigate these complicating factors along with potential decisions derived from the assessment information. An area of potential conflict arises when psychiatrists are attempting to fulfill roles of both administrator (caretaker) and psychotherapist, and yet have not clearly defined these roles either for themselves or for their patients. The resulting ambiguity may cause the patient to feel defensive and resistant and the psychiatrist to feel that the patient is not living up to the therapist's expectations. Elaboration of a specific trait or need in the patient cannot resolve this conflict, but must occur in the context of interactions between the therapist and the patient. A standard psychological evaluation investigating the internal structure of the patient does not address this issue.

A second possible problem area for clients referred in the midst of therapy can be the result of personal anxiety and discomfort on the therapist's part. Thus, issues such as countertransference and possibly the therapist's unreasonable expectations may be equally or even more important than looking at a patient's characteristics. If role ambiguity, countertransference, or unreasonable expectations are discovered, they must be elaborated and communicated in a sensitive manner.

When psychiatrists are acting in the role of physician, they and the psychologist may have different conceptual models for describing a patient's disorder. Whereas psychiatrists function primarily from a disease or medical model, psychologists may speak in terms of difficulties in living with people and society. In effectively communicating the results of psychological evaluations, examiners must bridge this conceptual difference. For example, a psychiatrist may ask whether a patient is schizophrenic, whereas a psychologist may not believe that the label *schizophrenia* is useful or even a scientifically valid concept. The larger issue, however, is that the psychiatrist is still faced with some practical decisions. In fact, the psychiatrist may share some of the same concerns regarding the term *schizophrenia,* but this conceptual issue may not be particularly relevant in dealing with the patient. Legal requirements or hospital policies might require that the patient be given a traditional diagnosis. The psychiatrist may also have to decide whether to give antipsychotic medication, electroconvulsive therapy, or psychotherapy. For a patient who is diagnosed as schizophrenic rather than brain-damaged or personality-disordered, then (given a hospital's current and economic policy considerations), the psychiatrist may decide on antipsychotic medication. An effective examiner should be able to see beyond possible conceptual differences and instead address practical considerations. A psychiatrist may refer a defensive patient who cannot or will not verbalize his or her concerns and ask whether this person is schizophrenic. Beyond this are factors such as the quality of the patient's thought processes and whether the person poses a danger to himself or herself or to others. Thus, the effective examiner must translate his or her findings into a conceptual model that is both understandable by a psychiatrist and useful from a task-oriented point of view.

General Medical Setting

It has been estimated that as many as two-thirds of patients seen by physicians have primarily psychosocial difficulties, and of those with clearly established medical diagnoses, between 25% to 50% have specifically psychological disorders in addition to

medical ones (Asaad, 2000; Katon & Walker, 1998; McLeod, Budd, & McClelland, 1997; Mostofsky & Barlow, 2000). Most of these psychological difficulties are neither diagnosed nor referred for treatment (American Journal of Managed Care, 1999; Borus, Howes, Devins, & Rosenberg, 1988; Mostofsky & Barlow, 2000). In addition, many traditionally "medical" disorders such as coronary heart disease, asthma, allergies, rheumatoid arthritis, ulcers, and headaches have been found to possess a significant psychosocial component (Groth-Marnat & Edkins, 1996; Pruit, Klapow, Epping-Jordan, & Dresselhaus, 1999). Not only are psychological factors related to disease, of equal importance, they are related to the development and maintenance of health. In addition, the treatment and prevention of psychosocial aspects of "medical" complaints have been demonstrated to be cost-effective for areas such as preparation for surgery, smoking cessation, rehabilitation of chronic pain patients, obesity, interventions for coronary heart disease, and patients who are somatizing psychosocial difficulties (Chiles, Lambert, & Hatch, 1999; Groth-Marnat & Edkins, 1996; Groth-Marnat, Edkins, & Schumaker, 1995; Sobel, 2000). A complete approach to the patient, then, involves an awareness of the interaction between physical, psychological, and social variables (Pruit et al., 1999; G. Schwartz, 1982). Thus, psychologists have the potential to make an extremely important contribution. To adequately work in general medical settings, psychologists must become familiar with medical descriptions, which often means learning a complex and extensive vocabulary. Another issue is that, even though physicians often draw information from several sources to aid in decision making, they must take ultimate responsibility for their decisions.

The most frequent situations in which physicians might use the services of a psychologist involve the presence of an underlying psychological disorder, possible emotional factors associated with medical complaints, assessment for neuropsychological deficit, psychological treatment for chronic pain, and the treatment of chemical dependency (Bamgbose et al., 1980; Groth-Marnat, 1988; Maruish, 2002). Although a medical exam may not suggest any physical basis for the patient's complaints, the physician still has to devise some form of treatment or at least an appropriate referral. This is crucial in that a significant portion of patients referred to physicians do not have any detectable physical difficulties and their central complaint is likely to be psychological (Asaad, 2000; Pruit et al., 1999; Maruish, 2002; Mostofsky & Barlow, 2000). The psychologist can then elaborate and specify how a patient can be treated for possible psychosocial difficulties (DeGood, Crawford, & Jongsma, 1999; Wickramasekera, 1995a, 1995b). This may require using not only the standard assessment instruments but also more specialized ones such as the Millon Behavioral Health Inventory or the Millon Behavioral Medicine Diagnostic (Bockian, Meagher, & Millon, 2000; Maruish, 2000; Millon, 1997).

Another area that has greatly increased in importance is the psychological assessment of a patient's neuropsychological status (see Chapter 12). Whereas physicians attempt to detect physical lesions in the nervous system, the neuropsychologist has traditionally been more concerned with the status of higher cortical functions. Another way of stating this: Physicians evaluate how the *brain* is functioning, whereas the neuropsychologist evaluates how the *person* is functioning as a result of possible brain abnormalities. The typical areas of assessment focus primarily on the presence of possible intellectual deterioration in areas such as memory, sequencing, abstract reasoning, spatial organization, or executive abilities (Groth-Marnat, 2000b). Such referrals, or at

least screening for neuropsychological deficit, typically account for approximately a third of all psychological referrals in psychiatric and medical settings. In the past, neuropsychologists have been asked to help determine whether a patient's complaints were "functional" or "organic." The focus now is more on whether the person has neuropsychological deficits that may contribute to or account for observed behavioral difficulties than on either/or distinctions (Loenberger, 1989). Physicians often want to know whether a test profile suggests a specific diagnosis, particularly malingering, conversion disorder, hypochondriasis, organic brain syndrome, or depression with pseudoneurological features. Further issues that neuropsychologists often address include the nature and extent of identified lesions, localization of lesions, emotional status of neurologically impaired patients, extent of disability, and suggestions for treatment planning such as recommendations for cognitive rehabilitation, vocational training, and readjustment to family and friends (Lemsky, 2000).

A physician might also request a psychologist to conduct a presurgical evaluation to assess the likelihood of a serious stress reaction to surgery. Finally, physicians—particularly pediatricians—are often concerned with detecting early signs of serious psychological disorder, which may have been brought to their attention by parents, other family members, or teachers. In such situations, the psychologist's evaluation should assess not only the patient's present psychological condition, but also the contributing factors in his or her environment, and should provide a prediction of the patient's status during the next few months or years. When the patient's current condition, current environment, and future prospects have been evaluated, the examiner can then recommend the next phase in the intervention process. A psychologist may also consult with physicians to assist them in effectively discussing the results of an examination with the patient or the patient's family.

Legal Context

During the past 15 years, the use of psychologists in legal settings has become more prevalent, important, and accepted (see Otto & Heilburn, 2002). Psychologists might be called in at any stage of legal decision making. During the investigation stage, they might be consulted to assess the reliability of a witness or to help evaluate the quality of information by a witness. The prosecuting attorney might also need to have a psychologist evaluate the quality of another mental health professional's report, evaluate the accused person's competency, or help determine the specifics of a crime. A defense attorney might use a psychologist to help in supporting an insanity plea, to help in jury selection, or to document that brain damage has occurred. A judge might use a psychologist's report as one of a number of factors to help determine a sentence, a penal officer might wish consultation to help determine the type of confinement or level of dangerousness, or a parole officer might need assistance to help plan a rehabilitation program. Even though a psychologist might write a legal report, he or she is likely to actually appear in court in only about one in every ten cases.

The increasing use and acceptance of psychologists in legal contexts have resulted in a gradual clarification of their roles (Blau, 1998; Otto & Heilburn, 2002), as well as a proliferation of forensic specific assessment instruments (Heilburn, 2001). However, acclimatizing to the courtroom environment is often difficult because of the quite different

roles between courtroom and clinic, as well as the need to become familiar with specialized legal terms such as *diminished capacity* and *insanity.* In addition, many attorneys are familiar with the same professional literature that psychologists read and may use this information to discredit a psychologist's qualifications, methods of assessment, or conclusions (Faust, Ziskin, & Hiers, 1991; Ziskin & Faust, 1995). Psychologists are also required to become increasingly sophisticated in their evaluation of possible malingering and deception (see R. Rogers, 1997 for guidelines).

Each psychologist appearing in court must have his or her qualifications approved. Important areas of consideration are the presence of clinical expertise in treating specialty disorders and relevant publication credits. Evaluation of legal work by psychologists indicates they are generally viewed favorably by the courts and may have reached parity with psychiatrists (Sales & Miller, 1994).

As outlined by the American Board of Forensic Psychology, the practice of forensic psychology includes training/consultation with legal practitioners, evaluation of populations likely to encounter the legal system, and the translation of relevant technical psychological knowledge into usable information. Psychologists are used most frequently in child custody cases, competency of a person to dispose of property, juvenile commitment, and personal injury suits in which the psychologist documents the nature and extent of the litigant's suffering or disability (stress, anxiety, cognitive deficit). In contrast, psychiatrists are far more likely to be used in assessing a person's competency to stand trial, degree of criminal responsibility, and the presence of mental defectiveness. Although psychologists can testify in these cases, physicians need to sign any commitment certificates and are, therefore, more likely to be used.

An essential requirement when working in the legal context is for psychologists to modify their language. Many legal terms have exact and specific meanings that, if misunderstood, could lead to extremely negative consequences. Words such as *incompetent, insane,* or *reasonable certainty* may vary in different judicial systems or from state to state. Psychologists must familiarize themselves with this terminology and the different nuances involved in its use. Psychologists may also be requested to explain in detail the meaning of their conclusions and how these conclusions were reached. Whereas attorneys rarely question the actual data that psychologists generate, the inferences and generalizability of these inferences are frequently placed under scrutiny or even attacked. Often this questioning can seem rude or downright hostile, but in most cases, attorneys are merely doing their best to defend their client. Proper legal protocol also requires that the psychologist answer questions directly rather than respond to the implications or underlying direction suggested by the questions. Furthermore, attorneys (or members of the jury) may not be trained in or appreciate the scientific method, which is the mainstay of a psychologist's background. In contrast, attorneys are trained in legal analysis and reasoning, which subjectively focus on the uniqueness of each case rather than on a comparison of the person to a statistically relevant normative group.

Two potentially problematic areas lie in evaluating insanity and evaluating competency. Even though physicians are more typically called on to testify in these areas, psychologists can also become involved. Although the insanity plea has received considerable publicity, very few people make the appeal; and, of those who do, few have it granted. It is usually difficult for an expert witness to evaluate such cases because of the problem of possible malingering to receive a lighter sentence and the possible

ambiguity of the term *insanity*. Usually a person is considered insane in accordance with the McNaughton Rule, which states that persons are not responsible if they did not know the nature and extent of their actions and if they cannot distinguish that what they did was wrong according to social norms. In some states, the ambiguity of the term is increased because defendants can be granted the insanity plea if it can be shown they were insane at the time of the incident. Other states include the clause of an "irresistible impulse" to the definition of insanity. Related to insanity is whether the defendant is competent to stand trial. *Competence* is usually defined as the person's ability to cooperate in a meaningful way with the attorney, understand the purpose of the proceedings, and understand the implications of the possible penalties. To increase the reliability and validity of competency and insanity evaluations, specialized assessment techniques have been developed; for example, the Competency Screening Test (Lipsitt, Lelos, & McGarry, 1971; Nottingham & Mattson, 1981) and the Rogers Criminal Responsibility Scales (R. Rogers, 1984).

The prediction of dangerousness has also been a problematic area. Because actual violent or self-destructive behavior is a relatively unusual behavior (low base rate) any cutoff criteria typically are going to produce a high number of false positives (Mulvey & Cauffman, 2001). Thus, people incorrectly identified may potentially be detained and understandably be upset. However, the negative result of failure to identify and take action against people who are potentially violent makes erring on the side of caution more acceptable. Attempts to use special scales on the MMPI (Overcontrolled Hostility Scale; Megargee & Mendelsohn, 1962) or a 4-3 code type (see Chapter 7) have not been found to be sufficiently accurate for individual decision making. However, significant improvements have been made in predicting dangerousness and reoffending by using actuarial strategies, formal ratings, and summed ratings, which include relevant information on developmental influences, possible events that lower thresholds, arrest record, life situation, and situational triggers such as interpersonal stress and substance intoxication (Monahan & Steadman, 2001; Monahan et al., 2000; Steadman et al., 2000). The legal/justice system is most likely to give weight to those individual assessment strategies that combine recidivism statistics, tests specifically designed to predict dangerousness, summed ratings, and double administrations of psychological tests to assess change over time. Clinical judgment combined with a single administration of tests is usually considered only mildly useful.

Psychologists are sometimes asked to help with child custody decisions. Guidelines for developing child custody evaluations and child protection evaluations have been developed by the American Psychological Association (APA) (Guidelines for Child Custody Evaluations in Divorce Proceedings, APA, 1994 and Guidelines for Psychological Evaluations in Child Protection Matters, APA, 1998). The central consideration is to determine which arrangement is in the child's best interest. Areas to be considered include the mental health of the parent, the quality of love and affection between the parent and child, the nature of the parent-child relationship, and the long-term effect of the different decisions on the child. Often, psychological evaluations are conducted on each member of the family using traditional testing instruments. Specific tests, such as the Bricklin Perceptual Scales (Bricklin, 1984), have also been developed.

A final, frequently requested service is to aid in the classification of inmates in correctional settings. One basic distinction is between merely managing the person versus attempting a program of rehabilitation. Important management considerations are levels

of suicide risk, appropriateness of dormitory versus a shared room, possible harassment from other inmates, or degree of dangerousness to others. Rehabilitation recommendations may need to consider the person's educational level, interests, skills, abilities, and personality characteristics related to employment.

Educational Context

Psychologists are frequently called on to assess children who are having difficulty in, or may need special placement in, the school system. The most important areas are evaluating the nature and extent of a child's learning difficulties, measuring intellectual strengths and weaknesses, assessing behavioral difficulties, creating an educational plan, estimating a child's responsiveness to intervention, and recommending changes in a child's program or placement (Sattler, 2001). Any educational plan should be sensitive to the interactions among a child's abilities, the child's personality, the characteristics of the teacher, and the needs and expectations of the parents.

A typical educational placement begins with a visit to the classroom for observation of a child's behavior under natural conditions. A valuable aspect of this is to observe the interaction between the teacher and child. Typically, any behavioral difficulty is closely linked with the child-teacher interaction. Sometimes the teacher's style of responding to a student can be as much a part of the problem as the student. Consequently, classroom observations can cause discomfort to teachers and should be handled sensitively.

Observing the child in a wider context is, in many ways, contrary to the tradition of individual testing. However, individual testing all too frequently provides a relatively limited and narrow range of information. If it is combined with a family or classroom assessment, additional crucial data may be collected, but there is also likely to be significant resistance. This resistance may result from legal or ethical restrictions regarding the scope of the services the school can provide or the demands that a psychologist can make on the student's parents. Often there is an initial focus on, and need to perceive, the student as a "problem child" or "identified patient." This may obscure larger, more complex, and yet more significant issues such as marital conflict, a disturbed teacher, misunderstandings between teacher and parents, or a conflict between the school principal and the parents. All or some of these individuals may have an investment in perceiving the student as the person with the problem rather than acknowledging that a disordered school system or significant marital turmoil may be responsible. An individually oriented assessment may be made with excellent interpretations, but unless wider contexts are considered, understood, and addressed, the assessment may very well be ineffective in solving both the individual difficulties and the larger organizational or interpersonal problems.

Most assessments of children in a school context include behavioral observations, a test of intellectual abilities such as the WISC-III, Stanford Binet, Woodcock-Johnson Psychoeducational Battery-III (Woodcock, McGrew, & Mather, 2001) or Kaufman Assessment Battery for Children (K-ABC; A. Kaufman & Kaufman, 1983), and tests of personality functioning. In the past, assessment of children's personality generally relied on projective techniques. However, many projective tests have been found to have inadequate psychometric properties and are time consuming to administer, score, and interpret. As a result, a wide variety of behavioral ratings instruments have begun

to replace the use of projective instruments (Kamphaus et al., 2000). These include the Achenbach Child Behavior Checklist (Achenbach, 1994), Conners' Parent Rating Scale-Revised (Conners, 1997), Conners' Teacher Rating Scale-Revised (Conners, Sitarenios, Parker, & Epstein, 1998), and the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992). A number of sound objective instruments, such as the Personality Inventory for Children (PIC; Wirt, Lachar, Klinedinst, & Seat, 1977), have also been developed. The inventory was designed along similar lines as the MMPI, but is completed by a child's parent. It produces 4 validity scales to detect faking and 12 clinical scales, such as Depression, Family Relations, Delinquency, Anxiety, and Hyperactivity. The scale was normed on 2,400 children, empirically developed, extensively researched, and has yielded good reliability. Assessment of adolescent personality can be effectively done with the MMPI-A or the Millon Adolescent Clinical Inventory (MACI; Millon, 1993). Additional well-designed scales that have become increasingly used are the Adaptive Behavior Inventory for Children (Mercer & Lewis, 1978), Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984), and the Wechsler Individual Achievement Test (WIAT; Psychological Corporation, 1992). Whereas the Wide Range Achievement Test (WRAT-III; Wilkinson, 1993) has been used quite extensively in the past, it is being used less frequently (Kamphaus et al., 2000) because of alternative available instruments (primarily the Wechsler Individual Achievement Test).

Any report written for an educational setting should focus not only on a child's weaknesses, but also on his or her strengths. Understanding a child's strengths can potentially be used to increase a child's self-esteem as well as to create change in a wide context. Recommendations should be realistic and practical. This can most effectively be developed when the clinician has a thorough understanding of relevant resources in the community, the school system, and the classroom environment. This understanding is particularly important because the quality and resources available between one school or school system and the next can vary tremendously. Recommendations typically specify which skills need to be learned, how these can be learned, a hierarchy of objectives, and possible techniques for reducing behaviors that make learning difficult. Recommendations for special education should be made only when a regular class would clearly not be equally beneficial. However, the recommendations are not the end product. They are beginning points that should be elaborated and modified depending on the initial results. Ideally, a psychological report should be followed up with continuous monitoring.

The assessment of children should be carried out in two phases. The first phase should assess the nature and quality of the child's learning environment. If the child is not exposed to adequate quality instruction, he or she cannot be expected to perform well. Thus, it must first be demonstrated that a child has not been learning even with appropriate instruction. The second phase involves a comprehensive assessment battery, which includes measures of intellectual abilities, academic skills, adaptive behavior, and screening out any biomedical disorders that might disrupt learning. Intellectual abilities might involve memory, spatial organization, abstract reasoning, and sequencing. Regardless of students' academic and intellectual abilities, they will not perform well unless they have relevant adaptive abilities, such as social skills, adequate motivation, and ability to control impulses. Assessing a child's values and attitudes toward education may be particularly important because they determine whether the student is

willing to use whatever resources he or she may have. Likewise, the person's level of personal efficacy helps to determine whether the person is able to perform behaviors leading toward attaining the goals he or she values. Physical difficulties that might interfere with learning include poor vision, poor hearing, hunger, malnutrition, or endocrine dysfunction.

The preceding considerations clearly place the assessment of children in educational settings into a far wider context than merely the interpretation of test scores. Relationships among the teacher, family, and student need to be assessed, along with the relative quality of the learning environment. Furthermore, the child's values, motivation, and sense of personal efficacy need to be taken into consideration, along with possible biomedical difficulties. Examiners need to become knowledgeable regarding the school and community resources as well as learn new instruments that have demonstrated relatively high levels of reliability and validity.

Psychological Clinic

In contrast to the medical, legal, and educational institutions where the psychologist typically serves as a consultant to the decision maker, the psychologist working in a psychological clinic is often the decision maker. A number of frequent types of referrals come into the psychological clinic. Perhaps the most common ones are individuals who are self-referred and are seeking relief from psychological turmoil. For most of these individuals, extensive psychological testing is not relevant and, in fact, may be contraindicated because the time spent in testing is usually time that could best be applied toward treatment. However, brief instruments targeted toward assessing client characteristics most relevant toward treatment planning can help to develop treatments that will speed the rate of treatment as well as optimize outcome (see Chapters 13 and 14). There may also be certain groups of self-referred clients about whom the psychologist may question whether the treatment available in a psychological clinic is appropriate. These clients can include persons with extensive medical problems, individuals with legal complications that need additional clarification, and persons who may require inpatient treatment. With these cases, it might be necessary to obtain additional information through psychological testing. However, the main purpose of the testing would be to aid in decision making rather than to serve as a direct source of help for the client.

Two other situations in which psychological assessment may be warranted involve children who are referred by their parents for school or behavioral problems and referrals from other decision makers. When referrals are made for poor school performance or behavioral problems involving legal complications, special precautions must be taken before testing. Primarily, the clinician must develop a complete understanding of the client's social network and the basis for the referral. This may include a history of previous attempts at treatment and a summary of the relationship among the parents, school, courts, and child. Usually a referral comes at the end of a long sequence of events, and it is important to obtain information regarding these events. After the basis of the referral has been clarified, the clinician may decide to have a meeting with different individuals who have become involved in the case, such as the school principal, previous therapists, probation officer, attorney, or teacher. This meeting may uncover myriad issues that require decisions, such as referral for family therapy, placement in a

special education program, a change in custody agreements between divorced parents, individual therapy of other members of the family, and a change in school. All of these may affect the relevance of, and approach to, testing, but these issues may not be apparent if the initial referral question is taken at face value. Sometimes psychologists are also confronted with referrals from other decision makers. For example, an attorney may want to know if an individual is competent to stand trial. Other referrals may involve a physician who wants to know whether a head-injured patient can readjust to his or her work environment or drive a car, or the physician may need to document changes in a patient's recovery.

So far, this discussion on the different settings in which psychological testing is used has focused on when to test and how to clarify the manner in which tests can be most helpful in making decisions. Several additional summary points must be stressed. As has been discussed previously, a referral source sometimes is unable to adequately formulate the referral question. In fact, the referral question is usually neither clear nor concise. It is the clinician's responsibility to look beyond the referral question and determine the basis for the referral in its widest scope. Thus, an understanding must be developed of the complexity of the client's social setting including interpersonal factors, family dynamics, and the sequence of events leading to the referral. In addition to clarifying the referral question, a second major point is that psychologists are responsible for developing knowledge about the setting for which they are writing their reports. This includes learning the proper language, the roles of the individuals working in the setting, the choices facing decision makers, and the philosophical and theoretical beliefs they adhere to. It is also important that clinicians understand the values underlying the setting and assess whether these values coincide with their own. For example, psychologists who do not believe in aversion therapy, capital punishment, or electroconvulsive therapy may come into conflict while working in certain settings. Psychologists, thus, should clearly understand how the information they give their referral source will be used. It is essential for them to appreciate that they have a significant responsibility, because decisions made regarding their clients, which are often based on assessment results, can frequently be major changing points in a client's life. If the possibility exists for the information to be used in a manner that conflicts with the clinician's value system, he or she should reconsider, clarify, or possibly change his or her relationship to the referral setting.

A final point is that clinicians should not allow themselves to be placed into the role of a "testing technician" or psychometrist. This role ultimately does a disservice to the client, the practitioner, and the profession. Clinicians should not merely administer, score, and interpret tests, but should also understand the total referral context in its broadest sense. This means they also take on the role of an expert who can integrate data from a variety of sources. Tests, by themselves, are limited in that they are not flexible or sophisticated enough to address themselves to complex referral questions. Levine (1981) writes:

> [The formal research on test validity is] not immediately relevant to the practical use of psychological tests. The question of the value of tests becomes not "Does this test correlate with a criterion?" or "Does the test accord with a nomological net?" but rather "Does the use of the test improve the success of the decision making process?"

> by making it either more efficient, less costly, more accurate, more rational, or more relevant. (p. 292)

All of these concerns are consistent with the emphasis on an examiner fulfilling the role of an expert clinician performing psychological assessment rather than a psychometrist acting as a technician.

ETHICAL PRACTICE OF ASSESSMENT

During the approximately 80 years that psychologists have been conducting formal assessment, a number of ethical guidelines have gradually evolved to ensure that appropriate professional relationships and procedures are developed and maintained. These guidelines have largely evolved through careful considerations of what constitutes ideal practice. Many of these considerations have been highlighted and refined because of difficulties surrounding assessment procedures. Criticism has been directed at the use of tests in inappropriate contexts, confidentiality, cultural bias, invasion of privacy, and the continued use of tests that are inadequately validated. This has resulted in restrictions on the use of certain tests, greater clarification within the profession regarding ethical standards, and increased skepticism from the public. To deal with these potential difficulties as well as conduct useful and accurate assessments, clinicians need to be aware of the ethical use of assessment tools. The American Educational Research Association (AERA) and other professional groups have published guidelines for examiners in their *Standards for Educational and Psychological Tests* (1999), *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association, 1992), and *Guidelines for Computer-Based Test Interpretations* (American Psychological Association, 1986). A special series in the *Journal of Personality Assessment* (Russ, 2001) also elaborates on ethical dilemmas found in training, medical, school, and forensic settings. The following section outlines the most important of these guidelines along with additional related issues.

Developing a Professional Relationship

Assessment should be conducted only in the context of a clearly defined professional relationship. This means that the nature, purpose, and conditions of the relationship are discussed and agreed on. Usually, the clinician provides relevant information, followed by the client's signed consent. Information conveyed to the client usually relates to the type and length of assessment, alternative procedures, details relating to appointments, the nature and limits of confidentiality, financial requirements, and additional general information that might be relevant to the unique context of an assessment (see Handelsman & Galvin, 1988 and Zuckerman's, 1997, *The Paper Office* for specific guidelines, formats, and forms for informed consent).

An important area to be aware of is the impact the quality of the relationship can have on both assessment results and the overall working relationship. It is the examiner's responsibility to recognize the possible influences he or she may exert on the client and to optimize the level of rapport. For example, enhanced rapport with older children (but not

younger ones) involving verbal reinforcement and friendly conversation has been shown to increase WISC-R scores by an average of 13 IQ points compared with an administration involving more neutral interactions (Feldman & Sullivan, 1971). This is a difference of nearly one full standard deviation. It has also been found that mildly disapproving comments such as "I thought you could do better than that" resulted in significantly lowered performance when compared with either neutral or approving ones (Witmer, Bornstein, & Dunham, 1971). In a review of 22 studies, Fuchs and Fuchs (1986) concluded that, on the average, IQ scores were four points higher when the examiner was familiar with the child being examined than when he or she was unfamiliar with the child. This trend was particularly pronounced for lower socioeconomic status children. Whereas there is little evidence (Lefkowitz & Fraser, 1980; Sattler, 1973a, 1973b; Sattler & Gwynne, 1982) to support the belief that African American students have lower performance when tested by European American examiners, it has been suggested that African American students are more responsive to tangible reinforcers (money, candy) than European American students, who generally respond better to verbal reinforcement (Schultz & Sherman, 1976). However, in a later study, Terrell, Taylor, and Terrell (1978) demonstrated that the main factor was the cultural relevance of the response. They found a remarkable 17.6-point increase in IQ scores when African American students were encouraged by African American examiners with culturally relevant comments such as "nice job, blood" or "good work, little brother." Thus, the rapport and feedback, especially if that feedback is culturally relevant, can significantly improve test performance. As a result, the feedback, and level of rapport should, as much as possible, be held constant from one test administration to the next.

A variable extensively investigated by Rosenthal and his colleagues is that a researcher/examiner's expectations can influence another person's level of performance (R. Rosenthal, 1966). This has been demonstrated with humans as well as laboratory rats. For example, when an experimenter was told to expect better performances from rats that were randomly selected from the same litter as "maze bright" (compared with "maze dull"), the descriptions of the rats' performance given by the experimenter conformed to the experimenter's expectations (R. Rosenthal & Fode, 1963). Despite criticisms that have been leveled at his studies and the finding that the magnitude of the effect was not as large as originally believed (Barber & Silver, 1968; Elashoff & Snow, 1971), Rosenthal maintains that an expectancy effect exists in some situations and suggests that the mechanisms are through minute nonverbal behaviors (H. Cooper & Rosenthal, 1980). He maintains that the typical effects on an individual's performance are usually small and subtle, and occur in some situations but not others. The obvious implication for clinicians is that they should continually question themselves regarding their expectations of clients and check to see whether they may in some way be communicating these expectations to their clients in a manner that confounds the results.

An additional factor that may affect the nature of the relationship between the client and the examiner is the client's relative emotional state. It is particularly important to assess the degree of the client's motivation and his or her overall level of anxiety. There may be times in which it would be advisable to discontinue testing because situational emotional states may significantly influence the results of the tests. At the very least, examiners should consider the possible effects of emotional factors and incorporate these into their interpretations. For example, it might be necessary to

increase the estimate of a client's optimal intellectual functioning if the client was extremely anxious during administration of an intelligence test.

A final consideration, which can potentially confound both the administration and, more commonly, the scoring of responses, is the degree to which the examiner likes the client and perceives him or her as warm and friendly. Several studies (Sattler, Hillix, & Neher, 1970; Sattler & Winget, 1970) have indicated that the more the examiner likes the client, the more likely he or she will be to score an ambiguous response in a direction favorable to the client. Higher scores can occur even on items in which the responses are not ambiguous (Egeland, 1969; Simon, 1969). Thus, "hard" scoring, as opposed to more lenient scoring, can occur at least in part because of the degree of subjective liking the examiner feels toward the client. Again, examiners should continually check themselves to assess whether their relationship with the client is interfering with the objectivity of the test administration and scoring.

Invasion of Privacy

One of the main difficulties examinees can encounter in relation to psychological tests is that the examiner might discover aspects of the client that he or she would rather keep secret. Also of concern is that this information may be used in ways that are not in the best interest of the client. The Office of Science and Technology (1967), in a report entitled *Privacy and Behavioral Research,* has defined *privacy* as "the right of the individual to decide for him/herself how much he will share with others his thoughts, feelings, and facts of his personal life" (p. 2). This right is considered to be "essential to insure dignity and freedom of self determination" (p. 2). The invasion of privacy issue usually becomes most controversial with personality tests because items relating to motivational, emotional, and attitudinal traits are sometimes disguised. Thus, persons may unknowingly reveal characteristics about themselves that they would rather keep private. Similarly, many persons consider their IQ scores to be highly personal.

Public concern over this issue culminated in an investigation by the Senate Subcommittee on Constitutional Rights and the House Subcommittee on Invasion of Privacy. Neither of these investigations found evidence of deliberate or widespread misuse of psychological tests (Brayfield, 1965). Dahlstrom (1969) has argued that public concern over the invasion of privacy is based on two basic issues. The first is that tests have been oversold to the public, with a resulting exaggeration of their scope and accuracy. The public is usually not aware of the limitations of test data and may often feel that tests are more capable of discovering hidden information than they actually are. The second misconception is that it is not necessarily wrong to obtain information about persons that they either are unaware of themselves or would rather keep private. The more important issue is how the information is used. Furthermore, the person who controls where or how this information is used is generally the client. The ethical code of the American Psychological Association (1992) specifically states that information derived by a psychologist from any source can be released only with the permission of the client. Although there may be exceptions regarding the rights of minors, or when clients are a danger to themselves or others, the ability to control the information is usually clearly defined as being held by the client. Thus, the public is often uneducated

regarding its rights and typically underestimates the power it has in determining how the test data will be used.

Despite ethical guidelines relating to invasion of privacy, dilemmas sometimes arise. For example, during personnel selection, applicants may feel pressured into revealing personal information on tests because they aspire to a certain position. Also, applicants may unknowingly reveal information because of subtle, nonobvious test questions, and, perhaps more important, they have no control over the inferences that examiners make about the test data. However, if a position requires careful screening and if serious negative consequences may result from poor selection, it is necessary to evaluate an individual as closely as possible. Thus, the use of testing for personnel in the police, delicate military positions, or important public duty overseas may warrant careful testing.

In a clinical setting, obtaining personal information regarding clients usually does not present problems. The agreement that the information be used to help clients develop new insights and change their behavior is generally clear and straightforward. However, should legal difficulties arise relating to areas such as child abuse, involuntary confinement, or situations in which clients may be a danger to themselves or others, ethical questions often arise. Usually, there are general guidelines regarding the manner and extent to which information should be disclosed. These are included in the American Psychological Association's *Ethical Principles of Psychologists and Code of Conduct* (1992), and test users are encouraged to familiarize themselves with these guidelines.

Adequate handling of the issue of an individual's right to privacy involves both a clear explanation of the relevance of the testing and obtaining informed consent. Examiners should always have a clear conception of the specific reasons for giving a test. Thus, if personnel are being selected based on their mechanical abilities, tests measuring areas such as general maladjustment should not ordinarily be administered. Examiners must continually evaluate whether a test, or series of tests, is valid for a particular purpose, and whether each set of scores has been properly interpreted in relation to a particular context. Furthermore, the general rationale for test selection should be provided in clear, straightforward language that can be understood by the client. Informed consent involves communicating not only the rationale for testing, but also the kinds of data obtained and the possible uses of the data. This does not mean the client should be shown the specific test subscales beforehand, but rather that the nature and intent of the test should be described in a general way. For example, if a client is told that a scale measures "sociability," this foreknowledge might alter the test's validity in that the client may answer questions based on popular, but quite possibly erroneous, stereotypes. Introducing the test format and intent in a simple, respectful, and forthright manner significantly reduces the chance that the client will perceive the testing situation as an invasion of privacy.

Inviolacy

Whereas concerns about invasion of privacy relate to the discovery and misuse of information that clients would rather keep secret, inviolacy involves the actual negative feelings created when clients are confronted with the test or test situation. Inviolacy is particularly relevant when clients are asked to discuss information they would rather not

think about. For example, the MMPI contains questions about many ordinarily taboo topics relating to sexual practices, toilet behavior, bodily functions, and personal beliefs about human nature. Such questions may produce anxiety by making the examinees more aware of deviant thoughts or repressed unpleasant memories. Many individuals obtain a certain degree of security and comfort by staying within familiar realms of thought. Even to be asked questions that may indicate the existence of unusual alternatives can serve as an anxiety-provoking challenge to personal rules and norms. This problem is somewhat related to the issue of invasion of privacy and it, too, requires one-to-one sensitivity as well as clear and accurate information about the assessment procedure.

Labeling and Restriction of Freedom

When individuals are given a medical diagnosis for physical ailments, the social stigmata are usually relatively mild. In contrast are the potentially damaging consequences of many psychiatric diagnoses. A major danger is the possibility of creating a self-fulfilling prophecy based on the expected roles associated with a specific label. Many of these expectations are communicated nonverbally and are typically beyond a person's immediate awareness (H. Cooper & Rosenthal, 1980; R. Rosenthal, 1966). Other self-fulfilling prophecies may be less subtle; for example, the person who is labeled as a chronic schizophrenic is, therefore, given only minimal treatment because chronic schizophrenics rarely respond and then do not improve, perhaps mainly because of having received suboptimal treatment. Another negative consequence of labeling is the social stigma attached to different disorders. Thus, largely because of the public's misconceptions of terms such as *schizophrenia,* labeled individuals may be socially avoided.

Just as labels imposed by others can have negative consequences, self-acceptance of labels can likewise be detrimental. Clients may use their labels to excuse or deny responsibility for their behavior. This is congruent with the medical model, which usually assumes that a "sick" person is the victim of an "invading disorder." Thus, in our society, "sick" persons are not considered to be responsible for their disorders. However, the acceptance of this model for behavioral problems may perpetuate behavioral disorders because persons see themselves as helpless, passive victims under the power of mental health "helpers" (Szasz, 1987). This sense of helplessness may serve to lower their ability to deal effectively with new stress. In contrast to this is the belief that clients require an increased sense of responsibility for their lives and actions to effectively change their behavior.

A final difficulty associated with labeling is that it may unnecessarily impose limitations on either an individual or a system by restricting progress and creativity. For example, an organization may conduct a study to determine the type of person who has been successful at a particular type of job and may then develop future selection criteria based on this study. This can result in the future selection of relatively homogeneous employees, which in turn could prevent the organization from changing and progressing. There may be a narrowing of the "talent pool," in which people with new and different ideas are never given a chance. In other words, what has been labeled as adaptive in the past may not be adaptive in the future. One alternative to this predicament is to look at future trends and develop selection criteria based on these trends. Furthermore, diversity might be incorporated into an organization so that different but compatible types

can be selected to work on similar projects. Thus, clinicians should be sensitive to the potential negative impact resulting from labeling by outside sources or by self-labeling, as well as to the possible limiting effects that labeling might have.

Competent Use of Assessment Instruments

To correctly administer and interpret psychological tests, an examiner must have proper training, which generally includes adequate graduate course work, combined with lengthy supervised experience (Turner et al., 2001). Clinicians should have a knowledge of tests and test limitations, and should be willing to accept responsibility for competent test use. Intensive training is particularly important for individually administered intelligence tests and for the majority of personality tests. Students who are taking or administering tests as part of a class requirement are not adequately trained to administer and interpret tests professionally. Thus, test results obtained by students have questionable validity, and they should clearly inform their subjects that the purpose of their testing is for training purposes only.

In addition to the preceding general guidelines for training, examiners should also acquire a number of specific skills (Moreland, Eyde, Robertson, Primoff, & Most, 1995; Turner et al., 2001). These include the ability to evaluate the technical strengths and limitations of a test, the selection of appropriate tests, knowledge of issues relating to the test's reliability and validity, and interpretation with diverse populations. Examiners need to be aware of the material in the test manual as well as relevant research both on the variable the test is measuring and the status of the test since its publication. This is particularly important with regard to newly developed subgroup norms and possible changes in the meaning of scales resulting from further research. After examiners evaluate the test itself, they must also be able to evaluate whether the purpose and context for which they would like to use it are appropriate. Sometimes an otherwise valid test can be used for purposes it was not intended for, resulting in either invalid or useless inferences based on the test data. Examiners must also be continually aware of, and sensitive to, conditions affecting the examinee's performance. These conditions may include expectations on the part of the examiner, minor variations from the standardized instructions, degree of rapport, mood of the examinee, or timing of the test administration in relation to an examinee's life changes. To help develop accurate conclusions, examiners should have a general knowledge of the diversity of human behavior. Different considerations and interpretive strategies may be necessary for various ethnic groups, sex, sexual orientation, or persons from different countries (see Sandoval, Frisby, Geisinger, Ramos-Grenier, & Scheuneman, 1999). A final consideration is that, if interns or technicians are administering the tests, an adequately trained psychologist should be available as a consultant or supervisor.

Specific data-based guidelines for test user qualifications have been developed by relevant professional organizations (American Psychological Association, 1988; Moreland et al., 1995; Turner et al., 2001) and these guidelines have been incorporated by most organizations selling psychological tests. Qualification forms request information regarding the purpose for using tests (counseling, research, personnel selection), area of professional expertise (marriage and family, social work, school), level of training (degrees, licenses), specific courses taken (descriptive statistics, career assessment), and

quality control over test use (test security, appropriate tailoring of interpretations). Persons completing the forms certify that they possess appropriate training and competencies and agree to adhere to ethical guidelines and legal regulations regarding test use.

In addition to being appropriately trained to use tests themselves, psychologists should not promote the use of psychological techniques by persons who are not qualified. This does not mean that all psychological tests should be used exclusively by psychologists because many tests are available to other professionals. However, psychologists should be generally aware of which tests require a high level of training (i.e., individually administered IQ tests) and those that are more generally available.

One of the important aspects of competent test use is that the tests should be used only for the purposes they were designed for. Typically, tests being extended beyond what they were designed for have been done in good faith and with good intentions. For example, an examiner might use a TAT or Rorschach as the primary means of inferring an individual's IQ. Similarly, the MMPI-2 or MCMI-III, which was designed to assess the extent of psychopathology in an individual, might be inappropriately used to assess a normal person's level of functioning. Although some conclusions can be drawn from the MMPI-2 relating to certain aspects of a normal person's functioning, or although IQ estimates based on projectives can be made, they should be considered extremely tentative. These tests were not designed for these purposes and, as a result, such inferences do not represent their strengths. A somewhat more serious misuse can occur when a test such as the MMPI-2 is used to screen applicants for some types of personnel selection. Results from MMPI-2-type tests are likely to be irrelevant for assessing most job-related skills. Of equal importance is that the information derived from the MMPI-2 is typically of a highly personal nature and, if used in many types of personnel selection, is likely to represent an invasion of privacy.

Interpretation and Use of Test Results

Interpreting test results should never be considered a simple, mechanical procedure. Accurate interpretation means not simply using norms and cutoff scores, but also taking into consideration unique characteristics of the person combined with relevant aspects of the test itself. Whereas tests themselves can be validated, the integration of information from a test battery is far more difficult to validate. It is not infrequent, for example, to have contradictions among different sources of data. It is up to the clinician to evaluate these contradictions to develop the most appropriate, accurate, and useful interpretations. If there are significant reservations regarding the test interpretation, this should be communicated, usually in the psychological report itself.

A further issue is that test norms and stimulus materials eventually become outdated. As a result, interpretations based on these tests may become inaccurate. This means that clinicians need to stay current on emerging research and new versions of tests. A rule of thumb is that if a clinician has not updated his or her test knowledge in the past 10 years, he or she is probably not practicing competently.

Part of remaining current means that psychologists should select their testing instruments, as well as any scoring and interpretation services, based on evidence related to the validity of the programs or tests. Part of this requires knowledge of the context of the situation (Turner et al., 2001). A well-validated test might have been found to be

quite valid in one context or population but not for another. Another issue that might have ethical considerations is conversion to or use of computerized or Internet-assisted technology (McMinn, Buchanan, Ellens, & Ryan, 1999; McMinn, Ellens, & Soref, 1999). Ultimately, any interpretations and recommendations regarding a client are the responsibility of the clinician. Placing a signature on a report means that the clinician is taking responsibility for the contents of the report. Indeed, an important difference between an actuarial formula or automated report and a practitioner is that the practitioner ultimately will be held accountable.

Communicating Test Results

Psychologists should ordinarily give feedback to the client and referral source regarding the results of assessment (Lewak & Hogan, 2003; also see Pope, 1992 for specific guidelines and responsibilities). This should be done using clear, everyday language. If the psychologist is not the person giving the feedback, this should be agreed on in advance and the psychologist should ensure that the person providing the feedback presents the information in a clear, competent manner. Unless the results are communicated effectively, the purpose of the assessment is not likely to be achieved. This involves understanding the needs and vocabulary of the referral source, client, and other persons, such as parents or teachers, who may be affected by the test results. Initially, there should be a clear explanation of the rationale for testing and the nature of the tests being administered. This may include the general type of conclusions that are drawn, the limitations of the test, and common misconceptions surrounding the test or test variable. If a child is being tested in an educational setting, a meeting should be arranged with the school psychologist, parents, teacher, and other relevant persons. Such an approach is crucial for IQ tests, which are more likely to be misinterpreted, than for achievement tests. Feedback of test results should be given in terms that are clear and understandable to the receiver. Descriptions are generally most meaningful when performance levels are clearly indicated along with behavioral references. For example, in giving IQ results to parents, it is only minimally relevant to say that their child has an IQ of 130 with relative strengths in spatial organization, even though this may be appropriate language for a formal psychological evaluation. A more effective description might be that their child is "currently functioning in the top 2% when compared with his or her peers and is particularly good at organizing nonverbal material such as piecing together puzzles, putting together a bicycle, or building a playhouse."

In providing effective feedback, the clinician should also consider the personal characteristics of the receiver, such as his or her general educational level, relative knowledge regarding psychological testing, and possible emotional response to the information. The emotional reaction is especially important when a client is learning about his or her personal strengths or shortcomings. Facilities should be available for additional counseling, if needed. If properly given, feedback is not merely informative but can actually serve to reduce symptomatic distress and enhance self-esteem (Armengol, Moes, Penney, & Sapienza, 2001; Finn & Tonsager, 1992; Lewak & Hogan, 2003). Thus, providing feedback can actually be part of the intervention process itself. Because psychological assessment is often requested as an aid in making important life decisions, the potential impact of the information should not be underestimated. Clinicians are usually

in positions of power, and with that comes responsibility in that the information that clients receive and the decisions they make based on this information is often with them for many years.

Maintenance of Test Security and Assessment Information

If test materials were widely available, it would be easy for persons to review the tests, learn the answers, and respond according to the impression they would like to make. Thus, the materials would lose their validity. This means that psychologists should make all reasonable efforts to ensure that test materials are secure. Specifically, all tests should be kept locked in a secure place and no untrained persons should be allowed to review them. Any copyrighted material should not be duplicated. In addition, raw data from tests should not ordinarily be released to clients or other persons who may misinterpret them. However, clients have a right to the reports themselves should they request them. They also have the right to have the information released to a person they designate but such a request should be in writing (see Zuckerman, 1997, *The Paper Office,* for forms and guidelines).

The security of assessment results should also be maintained. Ideally, this means that only designated persons (usually the referral source and client) should see the results unless the client provides a release of information. In reality, however, this ethical principal may sometimes be difficult to achieve. For example, many medical contexts expect most relevant treatment information (including psychological assessment results) to be kept in clients' charts. Typically, all members of the treatment team have access to the charts (Claassen & Lovitt, 2001). On one level, this represents a conflict between psychological and medical guidelines. On another level, it represents a conflict between benefit to the patient (that may be enhanced by the treatment team having access to his or her records) and patient autonomy (patient control over who and where information should go). Security of assessment results can also be compromised when a large number of organizations (insurance company, interacting rehabilitation provider, referral source) all want access to patient records. This has become a particular issue in the managed health care environment. The security of client records also becomes more tenuous when large interconnected databases potentially have access to patient data (McMinn, Buchanan, et al., 1999; McMinn, Ellens, et al., 1999).

Sometimes in legal contexts, the court or the opposing council may wish to see either raw data or the actual test materials. Under these conditions, the court should be informed that ethical guidelines as well as agreements made with the test distributor require that this information not be released to untrained persons. An acceptable alternative would be for the psychologist to designate a person with appropriate training to receive the information and explain the data or describe the test material (Tranel, 1994).

TEST BIAS AND USE WITH MINORITY GROUPS

Bias in testing refers to the presence of systematic error in the measurement of certain factors (e.g., academic potential, intelligence, psychopathology) among certain individuals or groups (Sandoval et al., 1999; Suzuki, Meller, & Ponterotto, 1996). The possible

presence of bias toward minority groups has resulted in one of the most controversial issues in psychological testing. More specifically, critics believe that psychological tests are heavily biased in favor of, and reflect the values of, European American, middle-class society. They argue that such tests cannot adequately assess intelligence or personality when applied to minority groups. Whereas the greatest controversy has arisen from the use of intelligence tests, the presence of cultural bias is also relevant in the use of personality testing. Over the past decade, discussion over bias has shifted from controversy over the nature and extent of bias to a more productive working through of how to make the most valid and equitable assessment based on current knowledge (see Dana, 2000; Handel & Ben-Porath, 2000; Sandoval et al., 1999).

The original controversies over test bias centered on determining whether tests are as valid for minority groups as for nonminorities. Undoubtedly, differences exist; however, the meaning that can be attributed to these differences has been strongly debated. A further question lies in identifying the cause of these differences. Some theorists have argued that the differences are primarily the result of environmental factors (Kamin, 1974; R. Rosenthal & Jacobson, 1968), whereas others stressed hereditary determination (A. R. Jensen, 1969, 1972; Rushton, 1994). Although the debate is not resolved, guidelines have been established by the Equal Employment Opportunity Commission (EEOC) for the use of psychological tests with minority groups in educational and industrial settings. The basic premise is that a screening device (psychological test) can have an adverse impact if it screens out a proportionally larger number of minorities than nonminorities. Furthermore, it is the responsibility of the employer to demonstrate that the procedure produces valid inferences for the specific purposes for which the employer would like to use it. If an industrial or educational organization does not follow the guidelines as defined by the EEOC (1970), the Office of Federal Contract Compliance has the direct power to cancel any government contract that the institution might have.

The degree of test validity when used with ethnic minorities is of central importance to the legal issues, research data, and guidelines for the individual clinician. If investigated from the perspective of content validity, popular individual intelligence tests appear on the surface to be culturally biased. This conclusion is based largely on early intuitive observations that many African American children and other minorities usually do not have the opportunity to learn the types of material contained in many of the test items. Thus, their lower scores may represent not a lack of intelligence, but merely a lack of familiarity with European American, middle-class culture. Critics of the tests point out that it would clearly be unfair to assess a European American's intelligence based on whether he or she knows what the *funky chicken* is or what *blood* means, or for that matter, to ask him or her the meaning of British terms such as *shilling* or *lorrie*. Low scores would simply measure an individual's relative unfamiliarity with a specific culture rather than his or her specific mental strengths. If this reasoning is used, many IQ and aptitude tests may appear on the surface to be culturally biased. However, studies in which researchers, to the best of their ability, eliminated biased test items or items that statistically discriminate between minorities and nonminorities have not been successful in altering overall test scores. In a representative study, 27 items were removed from the SAT that consistently differentiated minorities from nonminorities. This did little to change either the test takers' individual scores, or the differences

between the two groups (Flaugher & Schrader, 1978). Thus, the popular belief, based on a superficial appraisal of many psychological tests that biased items are responsible for test differences, does not appear to be supported by research.

Although test differences between minority and nonminority groups have frequently been found, the meaning and causes of these differences are open to debate. It has been demonstrated that African Americans consistently score lower than European Americans on the WISC-R (A. R. Jensen & Reynolds, 1982; Neisser et al., 1996), WAIS-R (A. Kaufman, McLean, & Reynolds, 1988), and SAT (Temp, 1971) although these differences may be decreasing (Vincent, 1991). However, when African Americans and European Americans of equal socioeconomic status were compared, the differences in IQ scores were greatly reduced (Loehelin, 1989). Likewise, the 5 *T*-score point differences found on MMPI scales F, 8, and 9 were also decreased or even insignificant when African Americans and European Americans were comparable in age, education, and other relevant demographic characteristics (Dahlstrom, Lachar, & Dahlstrom, 1986b; Timbrook & Graham, 1994). This suggests that many differences in test scores may result primarily from factors such as socioeconomic status rather than ethnicity.

Another consideration is the adequacy of the predictive validity of various tests when used with minority groups. Because one of the main purposes of these tests is to predict later performance, it is essential to evaluate the extent to which the scores in fact adequately predict areas such as a minority's performance in college. A representative group of studies indicates that the SAT actually overpredicts how well minorities will perform in college (A. R. Jensen, 1984; Kallingal, 1971; Pfeifer & Sedlacek, 1971; Reynolds, 1986). Furthermore, both the WISC and the WISC-R are equally as effective in predicting the academic achievement of both African Americans and European Americans in primary and secondary schools (Neisser et al., 1996; Reynolds & Hartlage, 1979). In actually working with minority groups, however, it is important to become familiar with different subgroup norms and to know the confidence with which predictions can be made based on the scores of these subgroups.

The preceding discussion of content and predictive validity represents the traditional defense of psychological tests. For many individuals, these defenses are still not sufficient. The two main choices, then, are either to outlaw all psychological tests for minority groups or to develop more appropriate psychological assessment approaches. A half-serious attempt toward a more appropriate measuring device is the Dove Counterbalance General Intelligence Test (Dove, 1968). It has since become referred to as the *Chitling Test* and includes items relevant for an African American inner-city culture, such as "A *handkerchief head* is: (a) a cool cat, (b) a porter, (c) an Uncle Tom, (d) a haddi, (e) a preacher." A similar attempt by R. Williams (1974) is his development of the Black Intelligence Test of Cultural Homogeneity (BITCH). Although neither test has been standardized and validated, both contain vocabulary words and experiences with which most African American children would be familiar but with which European American children would be unfamiliar.

A number of additional tests have been developed with the partial intent of using them in the assessment of ethnic minorities. These tend to emphasize nonverbal tasks and include the Leiter International Performance Scale, Peabody Picture Vocabulary Test-III, General Abilities Measure for Adults, Raven's Progressive Matrices, and the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998; McCallum,

Bracken, & Wasserman, 2001). Some of these tests have been found to have minimal cultural bias (see A. Kaufman & Lichtenberger, 2002). In addition, the Kaufman Assessment Battery for Children (K-ABC; A. Kaufman & Kaufman, 1983) demonstrates minimal cultural bias. Mean IQ scores for European Americans, African Americans, and Hispanics are relatively close; and there is some evidence that reliability and concurrent validity is comparable for different ethnic populations (A. Kaufman & Kaufman, 1983). The test is based on empirical developments in cognitive psychology and has a good record of reliability and validity.

The System of Multicultural Pluralistic Assessment (SOMPA; Mercer, 1979; Mercer & Lewis, 1978) provides an alternative and more complex method of evaluating minorities by using traditional assessment tools but correcting the bias involved with these tools. The assumption underlying this approach is that all cultural groups have the same average potential and any adequate assessment device should accurately test this potential for a particular individual. One of its primary goals is to differentiate between members of minorities who have been incorrectly labeled mentally retarded because of test bias and those who are in fact mentally retarded. The SOMPA method involves medical, social system, and pluralistic components. The *medical* component assesses whether students have any physical disorders that may be interfering with their level of performance. This assessment includes tests of hearing, vision, and motor function. The rationale for the medically oriented assessment is that children from lower socioeconomic groups are both more likely to have medical difficulties because of their harsher environment and less likely to obtain treatment for these difficulties because of financial constraints. The *social system* component uses traditional assessment tools, such as the WISC-R, to measure whether the student is functioning at a level consistent with social norms. The problem with this component is that it provides a narrow definition of successful functioning because the criteria are based on the dominant culture's definition of success. Thus, the final *pluralistic* component attempts to correct for the narrow approach in the social system component by evaluating an individual's test scores against a culturally similar group, thereby, it is hoped, adjusting for such variables as socioeconomic status and cultural background. Thus, comparisons are made between performances within a specific subgroup, rather than with the performance, values, and criteria of the dominant culture. The resulting adjusted scores are referred to as an individual's *Estimated Learning Potentials* (ELPs).

SOMPA has had a number of critics, most of whom argue that the criterion for judging it should be the adequacy with which it can predict school performance (D. Johnson & Danley, 1981; Oakland, 1980). Studies indicate that, whereas WISC-R scores correlate at a level of .60 with grade point average, SOMPA scores have a correlation of only .40 (Oakland, 1980). ELPs have also been found to have lower correlations with other forms of achievement than traditional IQ measures (Wurtz, Sewell, & Manni, 1985), and it is difficult to relate ELP results to specific applications in the classroom (Brooks & Hosie, 1984). Mercer refutes these criticisms by pointing out that her intent was not so much to predict school performance as to identify students who have been falsely classified as mentally retarded. Proponents of SOMPA have been so persuasive that it has been adopted by several states. Many people hoped that SOMPA would create more accurate labeling of mentally retarded students. However, students who are now labeled *normal* through the SOMPA approach, but were previously labeled *mentally retarded* or *learning*

disabled, might still require some additional form of special instruction. In fact, reclassifying students as normal through a calculation of ELPs may bar access of these students from special educational services. In addition, studies indicate that a high proportion of students classified as mentally retarded using the SOMPA are still likely to be minorities (Heflinger, Cook, & Thackrey, 1987) and that scores may be biased in favor of urban children, regardless of their ethnicity (Taylor, Sternberg, & Partenio, 1986). Because of these difficulties, SOMPA has probably not achieved its goal of equalizing educational opportunities for ethnic minority children, and thus should be used with caution for individual educational decision making.

As is true for ability tests and tests of scholastic aptitude, personality tests have the potential to be biased. The main research in this area has been performed on the MMPI and has consistently indicated that minority groups do score differently than do nonminorities. In general, African Americans scored higher than European Americans on scales F, 8, and 9 (Green & Kelly, 1988; Gynther & Green, 1980), but this pattern was not consistent across all populations (Greene, 1987, 1991, 2000). Even if consistent score differences were found, this does not mean these differences will be of sufficient magnitude to alter a clinician's interpretations, nor does it mean that predictions based on empirical criteria will be different. Studies using empirical criteria for prediction indicate that the MMPI does not result in greater descriptive accuracy for European Americans than African Americans (Elion & Megargee, 1975; Green & Kelly, 1988). In a review of MMPI performance for Asian Americans, African Americans, Hispanics, and Native Americans, Greene (1987) concluded that "the failure to find a consistent pattern of scale differences between any two ethnic groups in any publication suggests that it is premature to begin to develop new norms for ethnic groups" (p. 509). What seems to affect MMPI profiles more than ethnicity are moderator variables such as socioeconomic status, intelligence, and education. Furthermore, the existing differences may result from true differences in behavior and personality caused by the greater stresses often encountered by minorities. J. Graham (1987) suggests that, when MMPI scores are deviant, the clinician should tentatively accept these scores but make special efforts to explore the person's life situation and level of adjustment, and integrate this information with the test scores.

From this discussion, it should be obvious that the problems are both complicated and far from being resolved. Several general solutions have been suggested (see Suzuki et al., 1996). These include improving selection devices, developing different evaluation criteria, and changing social environments. Improving the use of selection devices involves paying continual attention to, and obtaining greater knowledge of, the meaning of different scores for different subgroups. This may include tailoring specific test scores to the types of decisions individuals may make in their lives. For example, African Americans typically achieve scores equal to European Americans on the verbal portion of the SAT, but their average scores on math are lower. This suggests that African American students have a greater development in their verbal skills than in their quantitative ones. This conclusion is further reflected by, and consistent with, the fact that African Americans are more likely to choose verbally oriented majors in college. Based on this, it may be more accurate to predict the future college performances of African Americans from their SAT verbal scores than from their SAT math scores.

Another approach to solving the problem of potential test bias is to develop different and more adequate criterion measures. For example, it has been found that WISC-R

scores correlate highly with teacher/classroom ratings for nonminorities, but not for minorities (Goldman & Hartig, 1976). This indicates that using teacher/classroom ratings as a criterion of academic achievement is not appropriate for minorities. In contrast, the WISC-R accurately predicts grade point average for both minorities and nonminorities, which suggests that grade point average is a better criterion measure. Perhaps of greater relevance is the actual prediction of an individual's career performance. Current test predictors for graduate schools (Law School Aptitude Test, Medical School Aptitude Test, etc.) give generally satisfactory predictions for later academic performance, but do not predict whether an individual will be, for example, a good attorney or physician. In fact, it has been shown that medical school grades themselves are not associated with later success as a physician (Loughmiller, Ellison, Taylor, & Price, 1970). This issue may become particularly pronounced in comparing the relative effectiveness of minorities and nonminorities when working in different cultural settings. For example, if a European American and a Hispanic attorney are placed in settings in which they work with Hispanics, it is probable that the Hispanic attorney would be more effective because he or she will have increased rapport and greater familiarity with the language and values of his or her clientele.

Another solution involves changing the social environment. Part of the rationale for emphasizing this approach is the belief held by many researchers that the differences in test scores between minorities and nonminorities are not because of test bias but rather because tests accurately reflect the effects of an unequal environment and unequal opportunities. Even though, in some situations, different minority norms and additional predictive studies on minority populations are necessary, the literature suggests that tests are not as biased as they have been accused of being. Removal of seemingly biased or discriminating SAT items still results in the same mean scores, the WISC-R provides accurate predictions of grade point average for both minorities and nonminorities, and the MMPI is usually equally as accurate for making behavioral predictions for African Americans as for European Americans. Tests themselves are not the problem but merely the means of establishing that, often, inequalities exist between ethnic groups. The goal should be to change unequal environments that can ideally increase a population's skills as measured by current tests of aptitude, IQ, and achievement. Whereas improving selection devices and developing different criterion measures are still important, future efforts should also stress more equal access to educational and career opportunities.

All of these solutions can give some direction to the profession in general, but it is the responsibility of individual clinicians to keep abreast of research relating to minority groups and to incorporate this knowledge into the interpretations they make of test scores. As Mercer (1979) has emphasized, test scores are neither valid nor invalid, but inferences by clinicians based on these scores are.

Probably the most important strategy is to maintain a flexible attitude combined with the use of alternative assessment strategies. This strategy might include a variety of nonverbal techniques, such as the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998), Raven's Progressive Matrices Test, or emphasis on the Performance Scales of the WAIS-III/WISC-III. In addition, dynamic testing shows promise in assessing the extent to which a client can benefit from various ongoing learning opportunities (learning potential; Grigorenko & Sternberg, 1998). Material beyond tests should also have a greater significance (teacher reports, discussions with parents, history, behavioral observations).

SELECTING PSYCHOLOGICAL TESTS

The most important factor in test selection is the extent to which the test is useful in answering the referral question. An assessment of neurological patients might use tests sensitive to cerebral deficit; depressed patients might be given the Beck Depression Inventory-II (A. T. Beck, Steer, & Brown, 1996); and pain patients might be given the McGill Pain Questionnaire (Melzack, 1975), Millon Behavioral Health Inventory (Millon, Green, & Meagher, 2000), or Illness Behavior Questionnaire (Pilowski, Spence, Cobb, & Katsikitis, 1984). Another important factor in test selection is a particular practitioner's training, experience, personal preferences, and familiarity with relevant literature. For example, a clinician who has received training in the MMPI-2 might be concerned about its ability to assess personality disorders and may rather choose to use an instrument such as the MCMI (Millon, 1994). Clinicians might also select an instrument because it has practical efficiency in terms of time and economy. Thus, they may wish to use simple behavioral predictions made by the client rather than use more expensive, time consuming, and, quite possibly, less accurate tests (Shrauger & Osberg, 1981). Computer-assisted instruments may also help to lower the costs of assessment primarily by reducing direct practitioner time and achieving greater speed for scoring and hypothesis generation.

The most frequently used assessment techniques are included in the following chapters. Contact details for the major psychological distributors, along with a partial listing of tests they carry, are listed in Appendix A on page 673. Various combinations of these tests typically constitute a core battery used by clinicians. However, it is often necessary to expand such a core battery depending on the specifics of the referral question. Table 2.1 provides a listing of the domains for assessment along with relevant tests. While some of these tests are thoroughly described in specific chapters dedicated to them, some may be relatively unfamiliar and practitioners should obtain additional information on them. Various sources are available for information about these and other tests. Such sources can provide important information for deciding whether to obtain the tests and incorporate them into a battery. Probably the most useful is the *Mental Measurements Yearbook,* which contains a collection of critical test reviews that include evaluations of the meaning of the available research on each test. The 13th *Mental Measurements Yearbook* was published in 1998 (Impara & Plake, 1998), but it may be necessary to consult previous editions as not all tests are re-reviewed in each new edition. The reviews are available in book form as well as online (*Mental Measurement Database*; see www.unl.edu/buros/catalog.html). *Tests in Print V* (L. Murphy et al., 1999) is associated with the *Mental Measurements Yearbook* but, rather than focusing on evaluating tests, lists information on each test such as its title, population it was designed for, available subtests, updating, author(s), and publisher. A further listing, description, and evaluation of tests can be found in Maddox (1997) *Tests: A Comprehensive Reference for Assessment in Psychology, Education, and Business,* which provides descriptive information on more than 3,500 tests. Practitioners interested in obtaining information on rating scales and other measures used in clinical practice might consult *Measures for Clinical Practice: A Sourcebook* (Corcoran, 2000). Neuropsychological tests are reviewed in the preceding resources as well as in Lezak's (1995) *Neuropsychological Assessment,* Spreen and Strauss's (1998) *A Compendium of Neuropsychological Tests,* and

Table 2.1 Assessment instruments relevant for specific response domains

- Cognitive functioning
  - General functioning
    - Mental Status Examination
    - Mini-Mental Status Examination (MMSE)
  - Intellectual functioning
    - Wechsler Adult Intelligence Scale-III
    - Wechsler Intelligence Scale for Children-III
    - Stanford-Binet (4th ed.)
    - Kaufman Assessment Battery for Children
    - Woodcock-Johnson Psychoeducational Battery-III
  - Memory functions
    - Wechsler Memory Scale-III
    - Rey Auditory Verbal Learning Test
    - California Verbal Learning Test
    - Benton Visual Retention Test
  - Visuoconstructive abilities
    - Bender Visual Motor Gestalt Test
    - Drawing tests
  - Content of thought processes
    - Thematic Apperception Test
    - Children's Apperception Test
- Emotional functioning and level of psychopathology
  - General patterns and severity
    - Minnesota Multiphasic Personality Inventory-2
    - Millon Clinical Multiaxial Inventory-III
    - Millon Adolescent Personality Inventory
    - Rorschach
    - Symptom Checklist 90-Revised
    - Brief Symptom Inventory
    - Personality Inventory for Children
  - Depression
    - Beck Depression Inventory
    - Hamilton Rating Scale for Depression
    - Children's Depression Inventory
  - Anxiety
    - State-Trait Anxiety Inventory
    - Fear Survey Schedule
    - Anxiety Disorders Interview Schedule
  - Sexual disturbance
    - Derogatis Sexual Functioning Inventory
  - Marital/family disturbance
    - Dyadic Adjustment Scale
    - Family Environment Scale
    - Marital Satisfaction Inventory
    - Draw a Family/Kinetic Family Drawing

(continued)

Table 2.1 *(Continued)*

- Interpersonal patterns
  - California Psychological Inventory
  - Rathus Assertiveness Schedule
  - Therapeutic Reactance Scale
- General personality measures
  - Sixteen Personality Factors
  - NEO-PI-R
  - Myers Briggs Type Indicator
  - Adjective Checklist
  - Taylor Johnson Temperament Analysis
  - Sentence completion tests
- Academic/school adjustment
  - Achenbach Child Behavior Checklist
  - Vineland Social Maturity Scale
  - Connors Behavior Rating Scales
  - Kinetic School Drawing
- Academic achievement
  - Peabody Individual Achievement Test
  - Wide Range Achievement Test-III
  - Wechsler Individual Achievement Test
- Adaptive Level
  - AAMD Adaptive Behavior Scale
  - Vineland Adaptive Behavior Scale
- Vocational interests
  - Career Assessment Inventory
  - Kuder Occupational Interest Survey
  - Self-Directed Search
  - Strong Interest Inventory
- Alcohol abuse
  - Michigan Alcoholism Screening Test
  - Alcohol Use Inventory
- Diagnosis
  - Diagnostic Interview Schedule
  - Schedule for Affective Disorders and Schizophrenia
  - Structured Clinical Interview for DSM
  - Structured Interview for DSM Personality Disorders
  - Diagnostic Interview for Children and Adolescents
- Prognosis and risk
  - Suicide potential
    - Scale of Suicide Ideation
    - Beck Hopelessness Scale
  - Schizophrenia prognosis
    - Camberwell Family Interview

specialty journals in neuropsychology, particularly *Neuropsychology Review.* A careful review of the information included in these references will frequently answer questions clinicians might have related to a test's psychometric properties, usefulness, appropriateness for different populations, details for purchasing, and strengths and limitations. Most of the questions listed in Table 1.1 (see Chapter 1) can be answered by consulting the preceding resources.

An important and current trend in research and practice on psychological assessment is the use of tests to generate a treatment plan (Beutler, Clarkin, & Bongar, 2000; Jongsma & Peterson, 1995; Maruish, 1999). Indeed, a basic objective of psychological assessment is to provide useful information regarding the planning, implementation, and evaluation of treatment. With the increased specificity of both treatment and assessment, this goal is becoming possible. For example, oppositional, resistant clients have been found to have optimal treatment outcomes when either self-directed or paradoxical interventions have been used (Beutler, Clarkin, & Bongar, 2000; Beutler, Sandowicz, Fisher, & Albanese, 1996). In addition, a problem's severity has clear implications for the restrictiveness of treatment (inpatient, outpatient) as well as treatment duration and intensity. Thus, clinicians should not select tests based simply on their diagnostic accuracy or psychometric properties, but they should also be concerned with the functional utility of the tests in treatment planning. Accordingly, Chapter 14 presents a systematic, integrated approach to transforming assessment results into a series of clear treatment recommendations.

Two special concerns in selecting tests are faking and the use of short forms. In many situations, clinicians might be concerned that persons will either consciously or unconsciously provide inaccurate responses (Lanyon, 1997). Thus, these clinicians may want to be sure to include and pay particular attention to such tests as the MMPI-2, MCMI-III, and CPI, which have validity scales incorporated into them. Although controversial, many projective techniques may be resistant to attempts at faking. Concerns regarding the time required for assessment may cause examiners to consider selecting short forms of instruments such as the WAIS-III or WISC-III. Although many short forms for cognitive tests seem sufficiently valid for screening purposes, their use as substitutes for the longer forms is not acceptable (A. Kaufman, Kaufman, Balgopal, & McLean, 1996; A. Kaufman & Lichtenberger, 2002). Attempts to develop short forms for the longer objective personality tests such as the MMPI have not been successful and have been discouraged by experts in the field (J. Graham, 2000). However, future computerized applications which tailor items based on a client's previous responses (adaptive testing) may result in the development of shortened administrations with acceptable psychometric properties (Archer, Tirrell, & Elkins, 2001).

During the evaluation of single cases, such as in clinical diagnoses and counseling, clinicians do not usually use formal combinations of test scores. Rather, they rely on their past judgment, clinical experience, and theoretical background to interpret and integrate test scores. However, for personnel decisions, academic predictions, and some clinical decisions (recidivism rate, suicide risk), clinicians may be advised to use statistical formulas (Garb, 1998b). The two basic approaches for combining test results are multiple regression equations and multiple cutoff scores. Multiple regression equations are developed by correlating each test or subtest with a criterion. The higher the correlation, the greater is the weight in the equation. The correlation of the

entire battery with the criterion measure gives an indication of the battery's highest predictive validity. For example, high school achievement can be predicted with the following regression equation, which combines IQ and CPI subtests:

$$\begin{aligned}\text{Achievement} = {} & .786 + .195\ \text{Responsibility} + .44\ \text{Socialization} \\ & - .130\ \text{Good Impression} + .19\ \text{Achievement via Conformance} \\ & + .179\ \text{Achievement via Independence} + .279\ \text{IQ}\end{aligned}$$

This equation raises the correlation with GPA to .68 as compared with .60 when using IQ alone (Megargee, 1972). This correlation indicates that academic achievement is dependent not only on intellectual factors, but also on psychosocial ones, such as responsibility, socialization, achievement via independence, and achievement via conformance, all of which are measured by the CPI. The second strategy, multiple cutoff scores, involves developing an optimum cutoff for each test or subtest. If the person is above a certain specified score (i.e., above the brain-damaged or schizophrenic range), the score can be used to indicate the presence of a certain characteristic. Although not all tests have equations or cutoffs developed for them, the decision to include a test in a battery may depend in part on the presence of such formal extensions of the tests. In addition, many of the computer-assisted interpretive packages use various actuarial formulas (usually in combination with expert interpretations) to develop their interpretations.

COMPUTER-ASSISTED ASSESSMENT

During the past 30 years, computer-assisted assessment has grown exponentially. By 1990, 17% of practicing psychologists frequently used computer-generated narratives, with an additional 36% using them on an occasional basis (Spielberger & Piotrowski, 1990). By 1999, the number of psychologists stating that they used some form of computer-assisted testing had increased to 40% (McMinn, Buchanan, et al., 1999). More than 400 software packages are available, many of which are listed in various catalogues published and distributed by test suppliers. At present, computers are used mainly for their clerical efficiency in scoring and data storage and to generate interpretive reports. Future uses of computers are likely to include features such as innovative presentation of items (i.e., adaptive testing), networked norms, novel presentation of stimuli (i.e., virtual reality), psychophysiological monitoring, and artificial intelligence (Garb, 2000; Groth-Marnat, 2000a). Computing in mental health has included not only computer-assisted assessment but also computer interviews, computerized diagnosis, computer-aided instruction, direct treatment intervention, clinical consultation, and simulated psychiatric interviews (McMinn, Buchanan, et al., 1999).

Computer-assisted administration and interpretation in neuropsychology have seen a number of particular advances (see review by Kane & Kay, 1992). Batteries have been developed mainly in large organizational contexts (military, Federal Aviation Authority) and focused on specialized types of problems. For example, the Neurobehavioral Evaluation System is particularly sensitive to the impact of environmental toxins (Groth-Marnat, 1993), COGSCREEN has been used in the selection of airline pilots, and the military's UTCPAB was originally developed to assess the impact of

drugs in the workplace. The Cambridge Neuropsychological Test Automated Batteries (CANTAB) have been found to detect and locate brain damage including early signs of Alzheimer's, Parkinson's, and Huntington's disease (Fray, Robbins, & Sahakian, 1996). Despite these developments, they currently do not have the extensive validation studies associated with the more traditional tests such as the Halstead Reitan Neuropsychological Test Battery (Reitan & Wolfson, 1993). Although the computer-assisted programs show considerable promise, they are currently used less than the more familiar individually administered neuropsychological tests or test batteries (Camara et al., 2000).

Computer-assisted assessment has a number of advantages. Computers can save valuable professional time, potentially improve test-retest reliability, reduce possible tester bias, and reduce the cost to the consumer by improving efficiency (Butcher, Perry, & Atlis, 2000; Groth-Marnat, 1999). Even greater benefits may someday be realized by incorporating more complicated decision rules in interpretation, collecting data on response latency and key pressure, incorporating computer-based models of personality, tailoring future questions to a client based on past responses, and estimating the degree of certainty of various interpretations (Groth-Marnat, 2000a, 2000b).

In the past, computer-assisted assessment has resulted in considerable controversy within mental health publications (Faust & Ziskin, 1989; Groth-Marnat & Schumaker, 1989), the popular media (C. Hall, 1983), and professional publications outside the mental health area (Groth-Marnat, 1985). A primary issue has been untested reliability and validity. Research on reliability, however, has typically indicated that computerized administrations have generally excellent reliability that is at least equivalent to the paper-pencil versions (Campbell et al., 1999). In addition, computer-administered versus paper-pencil results for traditional tests have generally been found to result in negligible differences in scores (Finger & Ones, 1999). This supports the view that if a paper-and-pencil version of the test is valid, a computerized version will also have equal validity resulting from the comparability in scores.

A further issue is the validity of computer-based test interpretation. Butcher et al. (2000) concluded a narrative review on the validity of computer-based interpretations by stating that in the vast majority of computer-based interpretations, 60% of the interpretations were appropriate. Shorter to mid-length narratives were generally considered to have a higher proportion of valid interpretations when compared with longer ones. In addition, the narrative statements contained in the computer-based reports were comparable to the types of statements made by clinicians. While this generally supports the use of computer-based interpretations, the finding that 40% or more of interpretations were not considered accurate means that the computer-based reports should be carefully evaluated. Thus, cutting and pasting computerized narratives into reports, results in unacceptably high error rates. Indeed, 42% of psychologists surveyed felt this procedure raised ethical concerns (McMinn, Ellens, et al., 1999). The previous summary clearly emphasizes that computer-based reports should not be used to replace clinical judgment but should instead be used as an adjunct to provide possible interpretations that the clinician needs to verify.

One concern is that many software packages are available to persons who do not possess appropriate professional qualifications. Ideally, qualified persons should be those who meet the requirements for using psychological tests in general (Turner et al., 2001). The American Psychological Association (1986) has also attempted to clarify these

standards in their *Guidelines for Computer-Based Test Interpretation.* However, Krug's (1993) *Psychware Sourcebook* indicated that approximately a fifth of the programs could be sold to the general public. The American Psychological Association guidelines specify that users "have an understanding of psychological or educational measurement, validation problems, and test research" and that practitioners "will limit their use of computerized testing to techniques which they are familiar and competent to use" (American Psychological Association, 1986, p. 8). Users should also "be aware of the method used in generating the scores and interpretation and be able to evaluate its applicability to the purpose for which it will be used" (American Psychological Association, 1986, pp. 8–9).

The preceding difficulties associated with computer-assisted instruction suggest a number of guidelines for users (Groth-Marnat & Schumaker, 1989). First, practitioners should not blindly accept computer-based narrative statements, but rather should ensure, to the best of their ability, that the statements are both linked to empirically based research and placed in the context of the unique history and unique situation of the client. Computers have, among other benefits, the strong advantage of offering a wide variety of possible interpretations to the clinician, but these interpretations still need to be critically evaluated. Far greater research needs to be performed on both the meaning of computer-administered test scores and on the narrative interpretations based on these scores. The developers of software should also be encouraged to provide enough information in the manual to allow proper evaluation of the programs and should develop mechanisms to ensure the updating of obsolete programs.

RECOMMENDED READING

Blau, T. (1998). *The psychologist as expert witness* (2nd ed.). New York: Wiley.

Maruish, M. E. (Ed.). (2000). *Handbook of psychological testing in primary care settings.* Mahwah, NJ: Erlbaum.

Zuckerman, E. L., & Guyett, I. P. R. (1997). *The paper work office: The tools to make your small psychotherapy practice work ethically, legally, and profitably—forms, guidelines, and resources* (2nd ed.). Pittsburgh, PA: Three Wishes Press.

Chapter 3

THE ASSESSMENT INTERVIEW

Probably the single most important means of data collection during psychological evaluation is the assessment interview. Without interview data, most psychological tests are meaningless. The interview also provides potentially valuable information that may be otherwise unobtainable, such as behavioral observations, idiosyncratic features of the client, and the person's reaction to his or her current life situation. In addition, interviews are the primary means for developing rapport and can serve as a check against the meaning and validity of test results.

Sometimes an interview is mistakenly thought to be simply a conversation. In fact, the interview and conversation differ in many ways. An interview typically has a clear sequence and is organized around specific, relevant themes because it is meant to achieve defined goals. Unlike a normal conversation, the assessment interview may even require the interviewer and interviewee to discuss unpleasant facts and feelings. Its general objectives are to gather information that cannot easily be obtained through other means, establish a relationship that is conducive to obtaining the information, develop greater understanding in both the interviewer and interviewee regarding problem behavior, and provide direction and support in helping the interviewee deal with problem behaviors. The interviewer must not only direct and control the interaction to achieve specific goals, but also have knowledge about the areas to be covered in the interview.

A basic dimension of an interview is its degree of structure. Some interviews allow the participants to freely drift from one area to the next, whereas others are highly directive and goal oriented, often using structured ratings and checklists. The more unstructured formats offer flexibility, possibly high rapport, the ability to assess how clients organize their responses, and the potential to explore unique details of a client's history. Unstructured interviews, however, have received frequent criticism, resulting in widespread distrust of their reliability and validity. As a result, highly structured and semistructured interviews have been developed that provide sound psychometric qualities, the potential for use in research, and the capacity to be administered by less trained personnel.

Regardless of the degree of structure, any interview needs to accomplish specific goals, such as assessing the client's strengths, level of adjustment, the nature and history of the problem, diagnosis, and relevant personal and family history. Techniques for accomplishing these goals vary from one interviewer to the next. Most practitioners use at least some structured aids, such as intake forms that provide identifying data and basic elements of history. Obtaining information through direct questions on intake forms frees the clinician to investigate other aspects of the client in a more flexible, open-ended manner. Clinicians might also use a checklist to help ensure that they have

covered all relevant areas. Other clinicians continue the structured format throughout most of the interview by using one of the formally developed structured interviews, such as the Schedule for Affective Disorders and Schizophrenia (SADS) or Structured Clinical Interview for the *DSM-IV* (SCID).

HISTORY AND DEVELOPMENT

Early Developments

The earliest form of obtaining information from clients was through clinical interviewing. At first, these interviews were modeled after question-and-answer medical formats, but later, the influence of psychoanalytic theories resulted in a more open-ended, free-flowing style. Parallel to the appearance of the psychoanalytically oriented interview was the development of the more structured and goal-oriented mental status examination originally formulated by Adolf Meyer in 1902. The mental status examination assessed relevant areas of a client's current functioning, such as general appearance, behavior, thought processes, thought content, memory, attention, speech, insight, and judgment. Professionals also expressed early interest in the relationship between biographical data and the prediction of occupational success or prognosis for specific disorders.

Regardless of the style used, the interviews all had these common objectives: to obtain a psychological portrait of the person, to conceptualize what is causing the person's current difficulties, to make a diagnosis, and to formulate a treatment plan. The difficulty with unstructured interviews is that they were (and still are) considered to have questionable reliability, validity, and cost-effectiveness. The first standardized psychological tests were developed to overcome these limitations. Tests could be subjected to rigorous psychometric evaluation and were more economical because they required less face-to-face contact with the person(s) being evaluated.

Developments during the 1940s and 1950s

During the 1940s and 1950s, researchers and clinicians began conceptualizing and investigating the following critical dimensions of interviews:

1. Content versus process.
2. Goal orientation (problem solving) versus expressive elements.
3. Degree of directiveness.
4. Amount of structure.
5. The relative amount of activity expressed by the participants.

These issues have been the focus of numerous research studies. A representative and frequently cited study on interviewer style was reported by W. Snyder (1945), who found that a nondirective approach was most likely to create favorable changes and self-exploration in clients. In contrast, a directive style using persuasion, interpretation, and

interviewer judgments typically resulted in clients being defensive and resistant to expressing difficulties. Strupp (1958) investigated the experience-inexperience dimension and found, among other things, that experienced interviewers expressed more warmth, a greater level of activity, and a greater number of interpretations. Level of empathy did not alter, regardless of the interviewer's degree of experience. Further, representative studies include Porter's (1950) in-depth evaluation of the effects of different types of responses (evaluative, probing, reassuring) and R. Wagner's (1949) early review, which questioned the reliability and validity of employment interviews.

Developments during the 1960s

A considerable amount of research in the 1960s was stimulated by C. Rogers (1961), who emphasized understanding the proper interpersonal ingredients necessary for an optimal therapeutic relationship (warmth, positive regard, genuineness). Elaborating on Roger's ideas, Truax and Carkhuff (1967) developed a five-point scale to measure interviewer understanding of the client. This scale was used for research on interviewing, therapist training, and as support for a client-centered theoretical orientation. Additional research efforts were also directed toward listing and elaborating on different categories of interactions such as clarification, summarizing, and confrontation.

Other investigators conceptualized interviewing as an interactive system in which the participants simultaneously influenced each other (Matarazzo, 1965; Watzlawick, Beavin, & Jackson, 1966). This emphasis on an interactive, self-maintaining system became the core for most early and later formulations of family therapy. The 1960s also saw the development and formalization of behavioral assessment, primarily in the form of goal-directed interviews that focused on understanding current and past reinforcers as well as on establishing workable target behaviors. Proponents of behavioral assessment also developed formal rating instruments and self-reports for areas such as depression, assertiveness, and fear.

Some attempts were made at integrating different schools of thought into a coherent picture, such as Beier's (1966) conceptualization of unconscious processes being expressed through nonverbal behaviors that could then be subject to covert social reinforcement. However, the 1960s (and part of the 1970s) were mostly characterized by a splintering into different schools of conflicting and competing ideologies. For example, client-centered approaches emphasized the importance of staying with the client's self-exploration; behavioral interviews emphasized antecedents and consequences of behavior; and family therapy focused on interactive group processes. Parallel progress was made within each of these different schools and within different disciplines, but little effort was devoted to cross-fertilization and/or integration.

Throughout the 1950s and 1960s, child assessment was conducted primarily through interviews with parents. Direct interviews with the child were considered to be for therapeutic purposes rather than for assessment. Differential diagnosis was unusual; almost all children referred to psychiatric clinics were either undiagnosed or diagnosed as "adjustment reactions" (Rosen, Bahn, & Kramer, 1964). Early research by Lapouse and Monk (1958, 1964) using structured interviews, indicated that mothers were more likely to report overt behaviors that are bothersome to adults (thumb-sucking, temper tantrums), but children were more likely to reveal covert difficulties (fears, nightmares).

Somewhat later, P. Graham and Rutter (1968), using structured interviews of children (rather than a parent), found interrater agreement was high for global psychiatric impairment (.84); moderate for attentional deficit, motor behavior, and social relations (.61 to .64); and low for more covert difficulties such as depression, fears, and anxiety (.30).

Developments during the 1970s

Assessment with adults and children during the 1970s saw a further elaboration and development of the trends of the 1960s, as well as increased emphasis on structured interviews. The interest in structured interviews was fueled largely by criticisms about the poor reliability of psychiatric diagnosis. A typical structured interview would be completed by the interviewer either during or directly after the interview, and the data would be transformed into such scales as organicity, disorganization, or depression-anxiety.

Initial success with adult structured interviews (e.g., Present State Examination, Renard Diagnostic Interview) encouraged thinking regarding the further development of child-structured interviews both for global ratings and specific content areas. Child assessment became concerned not only with information derived from parents, but also with the child's own experience. There was a trend toward direct questioning of the child, greater emphasis on differential diagnosis, and the development of parallel versions of structured interviews for both the parent(s) and child.

Behavioral strategies of interviewing for both children and adults not only emphasized the interviewee's unique situation, but also provided a general listing of relevant areas for consideration. Kanfer and Grimm (1977) outlined the areas an interviewer should assess as:

1. Behavioral deficiencies.
2. Behavioral excesses.
3. Inappropriate environmental stimulus control.
4. Inappropriate self-generated stimulus.
5. Problem reinforcement contingencies.

In a similar categorization, Lazarus (1973) developed his BASIC-ID model, which describes a complete assessment as involving behaviors (B), affect (A), sensation (S), imagery (I), cognition (C), interpersonal relations (I), and need for pharmacological intervention/drugs (D).

Additional themes in the 1970s included interest in biographical data, online computer technology, and the training of interviewer skills. Specifically, efforts were made to integrate biographical data for predicting future behavior (suicide, dangerousness, prognosis for schizophrenia) and for inferring current traits. J. W. Johnson and Williams (1977) were instrumental in developing some of the earliest online computer technology to collect biographical data and to integrate it with test results. Although training programs were devised for interviewers, a central debate was whether interview skills could actually be significantly learned or improved (Wiens, 1976).

Whereas most reviews of the literature in the 1970s emphasized the advantages of a comprehensive structured format, family therapists were dealing with group processes

in which formal interview structure was typically deemphasized. Because most family therapists were observing fluid interactional processes, they needed to develop a vocabulary different from that used in traditional psychiatric diagnosis. In fact, *DSM* categories were usually considered irrelevant because they described static characteristics of individuals rather than ongoing group processes. Few, if any, structured formats were available to assess family relationships.

Developments during the 1980s

Many of the trends, concepts, and instruments developed in the 1960s and 1970s were further refined and adapted for the 1980s. One important effort was the adaptation of many instruments to the *DSM-III* (1980) and *DSM-III-R* (1987). In addition, the increased delineation of childhood disorders required greater knowledge related to differential diagnosis and greater demand for structured interviews as adjuncts to assessment. Many of the efforts were consistent with the use of specific diagnostic criteria along with a demand for efficiency, cost-effectiveness, and accountability. Despite concerns regarding computer-based interpretations (Groth-Marnat & Schumaker, 1989), some of these functions were beginning to be performed by specific computer programs. Because interviews were becoming increasingly structured, with the inclusion of scales and specific diagnostic strategies, the distinction between tests and interviews was becoming less clear. In some contexts, aspects of interviewing were even replaced with computer-requested and computer-integrated information and combined with simple programs to aid in diagnosis, such as DIANO III (Spitzer, Endicott, & Cohen, 1974) and CATEGO (Wing, Cooper, & Sartorius, 1974). During the mid- and late 1980s, most clinicians, particularly those working in large institutions, used a combination of structured interviews and open-ended unstructured approaches. Some research focused on the importance of the initial interview regarding clinical decision making and later therapeutic outcome (Hoge, Andrews, Robinson, & Hollett, 1988; Turk & Salovey, 1985). There was also a greater appreciation and integration of the work from different disciplines and from differing theoretical persuasions (Hersen, 1988). Finally, greater emphasis was placed on the impact and implications of culture and gender on the assessment process (L. Brown, 1990).

The 1990s and Beyond

Two of the defining features of psychology in the 1990s were managed health care and the controversy over the validity of repressed memories. Both of these issues had significant implications for interviewing. Managed health care emphasized the cost-effectiveness of providing health services; and for interviewing, this means developing the required information in the least amount of time. This may mean streamlining interviews by maximizing computer-derived information or paper-pencil forms. This brings up the larger issue of the extent to which practitioners need to spend face-to-face time with the client versus deriving information through other means. The development of single-session therapy (Hoyt, 1994) illustrates the potential brevity of information that might be required before making therapeutic interventions. There was also recognition that precise patient-treatment matching can optimize the treatment

and potentially the cost-effectiveness of psychosocial interventions (Antony & Barlow, 2002; Beutler & Clarkin, 1990; Beutler, Clarkin, & Bongar, 2000).

The controversy over repressed memories has forced interviewers to clarify the extent to which the information they derive from clients represents literal as opposed to narrative truth. Research has consistently indicated that client self-reports are reconstructions of events (Henry, Moffitt, Caspi, Langley, & Silva, 1994; Loftus, 1993) and are likely to be particularly questionable for retrospective reports of psychosocial variables (Henry et al., 1994). The even greater challenge to interviewers is to ensure that their interviewing style and method of questioning are not distorting the information derived from clients. This issue becomes intensely highlighted during interviews to investigate the possibility of childhood sexual abuse (see guidelines in S. White & Edelstein, 1991).

Further, continuing themes in the 1990s were the impact of gender and cultural issues and the further development of structured interviews. In some cases, the preceding issues have produced tension. For example, the greater demands for brief focused interventions contradict the emphasis of structured interviews on detailed and often time-consuming procedures. In addition, there has been greater clinical and political importance attached to detecting and treating childhood abuse; yet research and media coverage of recovered memories have suggested that some, if not many, of these memories are of questionable validity. The themes related to cost-effectiveness, patient-treatment matching, recovered memories, use of structured interviews, and cultural and gender issues are far from resolved and will continue to be important themes during this decade.

ISSUES RELATED TO RELIABILITY AND VALIDITY

Although the interview is not a standardized test, it is a means of collecting data and, as such, can and should be subjected to some of the same types of psychometric considerations as a formal test. This is important because interviews might introduce numerous sources of bias, particularly if the interviews are relatively unstructured. Reliability of interviewers is usually discussed in relation to interrater (interviewer) agreement. R. Wagner's (1949) early review of the literature found tremendous variation, ranging from .23 to .97 ($Mdn = .57$) for ratings of personal traits and −.20 to .85 ($Mdn. = .53$) for ratings of overall ability. Later reviews have generally found similar variations in interrater agreement (Arvey & Campion, 1982; L. Ulrich & Trumbo, 1965). The problem then becomes how to determine which ratings to trust and which to view with skepticism. Of particular relevance is why some interviewers focus on different areas and have different biases. A consistent finding is that, when interviewers were given narrow areas to assess and were trained in interviewer strategies, interrater reliability increased (Dougherty, Ebert, & Callender, 1986; Zedeck, Tziner, & Middlestadt, 1983). The consensus was that highly structured interviews were more reliable (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994). However, increased structure undermines one of the greatest strengths of interviews—their flexibility. In many situations, a free-form, open-ended approach may be the only way to obtain some types of information.

Research on interview validity has typically focused on sources of interviewer bias. For example, halo effects result from the tendency of an interviewer to develop a general impression of a person and then infer other seemingly related characteristics. For example, clients who are considered to express warmth may be seen as more competent or mentally healthy than they actually are. This clustering of characteristics may be incorrect, thereby producing distortions and exaggerations. Similarly, first impressions have been found to bias later judgments (W. Cooper, 1981). Confirmatory bias might occur when an interviewer makes an inference about a client and then directs the interview to elicit information that confirms the original inference. For example, a psychoanalytically-oriented interviewer might direct questions related to early childhood traumas, possibly incorrectly confirming traditional psychoanalytic explanations of current adult behaviors. Similar to halo effects is the finding that one specific outstanding characteristic (educational level, physical appearance, etc.) can lead an interviewer to judge other characteristics that he or she incorrectly believes are related to the outstanding one. For example, physical attractiveness has been found to create interviewer bias in job applicants (Gilmore, Beehr, & Love, 1986). In a clinical context, physical attractiveness may result in practitioners' either deemphasizing pathology or, on occasion, exaggerating pathology because of discomfort the interviewer may feel over his or her feelings of attraction (L. Brown, 1990). Interviewers also may focus incorrectly on explanations of behavior that emphasize traits rather than situational determinants (Ross, 1977). This error is particularly likely when the interpretation of interview data relies heavily on psychological tests, because tests, by their nature, conceptualize and emphasize static characteristics of the person rather than ongoing interactional processes.

In addition to the interviewer's perceptual and interactional biases, the interviewees may distort their responses. For example, they may present an overly favorable view of themselves, particularly if they are relatively naive regarding their motivations. Distortions are most likely found in sensitive areas such as sexual behavior. Some specific areas of distortions are represented by the finding that victims of automobile accidents typically exaggerated the amount of time they lost from work, 40% of respondents provided overestimates of their contributions to charity, and 17% of respondents reported their ages incorrectly (R. Kahn & Cannell, 1961). More extreme cases of falsification occur with outright (conscious) lies, delusions, confabulations, and lies by pathological (compulsive) liars that they partially believe themselves (Kerns, 1986). Inaccuracies based on retrospective accounts have been found to most likely occur related to psychosocial information (e.g., family conflict, onset of psychiatric symptoms) compared with variables such as change of residence, reading skill, height, and weight (B. Henry et al., 1994).

Reviews of interview validity, in which interviewer ratings were compared with outside criterion measures, have, like reliability measures, shown tremendous variability ranging from −.05 to +.75 (Arvey & Campion, 1982; Henry et al., 1994; Huffcutt & Arthur, 1994; J. Hunter & Hunter, 1984; L. Ulrich & Trumbo, 1965). One clear finding is that validity increases as the structure of the interview format increases (Huffcutt & Arthur, 1994; Marchese & Muchinsky, 1993). For example, a meta-analysis by Wiesner and Cronshaw (1988) found that unstructured interviews had validity coefficients of .20,

structuring the interview increased the validity to .63, and structured interviews by a panel using consensus ratings increased validity coefficients to a quite respectable .64. However, the validity seems to vary according to the type of variable that is being assessed. Situational employment interviews (asking the interviewee what he or she would do in a particular situation) had higher validities (.50) than interviews used to assess past job-related behavior (.39) or rate psychological qualities such as dependability (.29; McDaniel et al., 1994). It has also been found that interview accuracy increases more when interviewees are held accountable for the process they went through when coming to their decisions, compared to being held accountable for the accuracy of their predictions (procedural versus outcome accountability; Brtek & Motowidlo, 2002).

The previous brief review indicates that adding structure to interviews and paying close attention to the procedure by which decisions are made typically results in higher levels of validity. It also means that information derived from unstructured interviews should be treated cautiously and treated as tentative hypotheses that need to be supported by other means. Interviewers should also continually question the extent to which their particular style, attitudes, and expectations might be compromising interview validity. Given the difficulties related to unstructured formats, a variety of formal structured clinical interviews has been developed. Additional information on the reliability and validity of the most frequently used structured clinical interviews is provided in the last section of this chapter.

ASSETS AND LIMITATIONS

Both structured and unstructured interviews allow clinicians to place test results in a wider, more meaningful context. In addition, biographical information from interviews can be used to help predict future behaviors; what a person has done in the past is an excellent guide to what he or she is likely to continue doing in the future. Factors for predicting suicide risk, success in certain occupations, and prognosis for certain disorders can usually be most effectively accomplished by attending to biographical data rather than test scores. Because tests are almost always structured or "closed" situations, the unstructured or semistructured interview is typically the only time during the assessment process when the clinician can observe the client in an open, ambiguous situation. Observations can be made regarding how persons organize their responses, and inferences can be derived from subtle, nonverbal cues. These inferences can be followed up with further, more detailed questioning. This flexibility inherent in unstructured and semistructured interviews is frequently their strongest advantage over standardized tests. The focus during unstructured interviews is almost exclusively on the individual rather than on how that individual does or does not compare with a larger normative comparison group. Some types of information can be obtained only through this flexible, person-centered approach, which allows the interviewer to pay attention to idiosyncratic factors. In crisis situations when relatively rapid decisions need to be made, it can be impractical to take the time required to administer and interpret tests, leaving interviews and rapid screening devices as the only means of assessment. Finally, interviews allow clinicians to establish rapport and encourage client self-exploration. Rarely do

clients reveal themselves nor do they perform optimally on tests unless they first sense trust, openness, and a feeling of being understood.

The greatest difficulty with unstructured interviews is interviewer bias from perceptual and interactional processes such as the halo effect, confirmatory bias, and the primacy effect. This bias typically results in considerable variability for both reliability and validity as well as in difficulty comparing one subject with the next. One of the main reasons for diagnostic disagreement is variations in the information obtained (information variance) and variations in the criteria (criterion variance) used to conclude the presence or absence of a condition. In more concrete terms, this means that different practitioners develop and ask a wide variety of questions and apply standards for the presence of a condition, such as depression, in an inconsistent fashion. A further difficulty is the high cost of using trained interviewers for large-scale epidemiological studies.

Structured interviews have many distinct advantages over unstructured approaches. Because structured interviews have more psychometric precision, the results enable comparability between one case or population and the next. The standardized presentation allows for the development of reliable ratings, reduces information variance, and uses consistent diagnostic criteria (R. Rogers, 1995; Summerfeldt & Antony, 2002). In addition, the comprehensiveness of many structured interviews reduces the likelihood of missing a diagnosis or set of relevant symptomology. Partially because of these advantages, structured clinical interviews have progressed from being used primarily for research to use in a number of clinical settings. At issue, however, is the time required for structured interviews. The more recently developed computer-assisted programs offer a potential method of countering this difficulty (Epstein & Klinkenberg, 2001). In addition, instruments such as the Diagnostic Interview Schedule and Diagnostic Interview for Children and Adolescents have been designed for administration by lay interviewers, thereby reducing the time required by professionals.

Although structured interviews generally have higher psychometric properties than unstructured formats, they tend to overlook the idiosyncrasies and richness of the person. In many cases, these unique aspects may go undetected and yet may make a significant difference in interpreting test scores or making treatment recommendations. Although still somewhat controversial (Helzer & Robins, 1988), another criticism of many clinicians and researchers is that a highly structured approach may not create enough rapport for the client to feel sufficiently comfortable about revealing highly personal information. This is truer for the highly structured interviews, such as the Diagnostic Interview Schedule, than for a semistructured instrument, such as the Schedule for Affective Disorders and Schizophrenia, which includes an initial, relatively unstructured component. However, M. Rosenthal (1989) has noted that rapport with structured instruments can be enhanced through carefully educating the client as to the importance and procedures of these more structured approaches.

Although many of the structured interviews have demonstrated adequate reliability, studies relating to validity have primarily focused on the general level of impairment or simple discriminations between psychiatric and nonpsychiatric populations. There has been considerable controversy over what exactly is an acceptable outside criterion measure regarding the "true" diagnosis. In-depth studies of construct validity or incremental validity have yet to be performed. Furthermore, far more work needs to be done

on the treatment utility of structured interviews in areas such as prognosis, selection of treatment, and likely response to specific forms of pharmacological or psychotherapeutic interventions.

THE ASSESSMENT INTERVIEW AND CASE HISTORY

General Considerations

The previously mentioned historical and psychometric considerations indicate that no single correct way exists to conduct an unstructured or semistructured interview. Interviewer style is strongly influenced by theoretical orientation and by practical considerations. Persons strongly influenced by client-centered theories tend to be nondirective and avoid highly structured questions. This is consistent with the underlying belief that persons have the inner ability to change and organize their own behaviors. The goal of a client-centered interview, then, is to create the type of interpersonal relationship most likely to enhance this self-change. In contrast, a behavioral interview is more likely to be based on the assumption that change occurs because of specific external consequences. As a result, behavioral interviews are relatively structured because they are directed toward obtaining specific information that would help to design strategies based on altering external conditions. In addition, different interviewing styles and strategies work well with some clients but may be relatively ineffective with others.

A useful distinction is between a diagnostic interview and one that is more informal and exploratory. The goal of a diagnostic interview is to develop a specific diagnosis, usually based on the multiaxial *DSM-IV* model (see Othmer & Othmer, 1994; R. Rogers, 1995; Sommers-Flanagan & Sommers-Flanagan, 1999). This might follow a five-step process in which the clinician develops diagnostic clues, considers these in relation to diagnostic criteria, takes a psychiatric history, and, based on this information, develops a multiaxial diagnosis with corresponding estimates of prognosis (Othmer & Othmer, 1994). Such an interview is likely to be directive with a careful consideration of inclusion and exclusion criteria for different disorders. It is most likely to occur in a psychiatric or general medical setting. In contrast, many practitioners do not believe in the value of formal diagnosis and, accordingly, do not pursue a formal *DSM-IV* (1994) diagnosis. They might be more concerned with areas such as a client's coping style, social supports, family dynamics, or the nature of their disability. As such, their interviews might be less directive and more flexible. Again, neither style is right or wrong, but instead, may be appropriate and effective in one context (or client), whereas it is ineffective or inappropriate in another context.

Often, interviewers might wish to construct a semistructured interview format by listing in sequence the types of questions they would like to ask the person. To construct such a list, interviewers might consult Table 3.1 to note possibly relevant areas. Each of these areas might then be converted into specific questions. For example, the first few areas might be converted into the following series of questions:

- "What are some important concerns that you have?"
- "Could you describe the most important of these concerns?"

Table 3.1 Checklist for an assessment interview and case history

History of the Problem
- Description of the problem
- Initial onset
- Changes in frequency
- Antecedents/consequences
- Intensity and duration
- Previous treatment
- Attempts to solve
- Formal treatment

Family Background
- Socioeconomic level
- Parent's occupation(s)
- Emotional/medical history
- Married/separated/divorced
- Family constellation
- Cultural background
- Parent's current health
- Family relationships
- Urban/rural upbringing

Personal History

Infancy
- Developmental milestones
- Family atmosphere
- Amount of contact with parents
- Early medical history
- Toilet training

Early and Middle Childhood
- Adjustment to school
- Academic achievement
- Hobbies/activities/interests
- Peer relationships
- Relationship with parents
- Important life changes

Adolescence
- All areas listed for early and middle childhood
- Presence of acting out (legal, drugs, sexual)
- Early dating
- Reaction to puberty

Early and Middle Adulthood
- Career/occupational
- Interpersonal relationships
- Satisfaction with life goals
- Hobbies/interests/activities
- Marriage
- Medical/emotional history
- Relationship with parents
- Economic stability

Late Adulthood
- Medical history
- Ego integrity
- Reaction to declining abilities
- Economic stability

Miscellaneous
- Self-concept (like/dislike)
- Happiest/saddest memory
- Earliest memory
- Fears
- Somatic concerns (headaches, stomachaches, etc.)
- Events that create happiness/sadness
- Recurring/noteworthy dreams

- "When did the difficulty first begin?"
- "How often does it occur?"
- "Have there been any changes in how often it has occurred?"
- "What happens after the behavior(s) occurs?"

Because clients vary regarding their personal characteristics (age, educational level, degree of cooperation) and type of presenting problem (childhood difficulties, legal problems, psychosis), the questions necessarily need to vary from person to person. Furthermore, any series of questions should not be followed rigidly, but with a certain degree of flexibility, to allow exploring unique but relevant areas that arise during the interview.

Good interviewing is difficult to define, partly because different theoretical perspectives exist regarding clinician-client interaction. Furthermore, clinicians achieve successful interviews not so much by what they do or say, but by making sure they express the proper attitude. Whereas clinicians from alternative theoretical persuasions might differ regarding areas such as their degree of directiveness or the type of information they should obtain, they would all agree that certain aspects of the relationship are essential (Patterson, 1989). These include the interviewer's expression of sincerity, acceptance, understanding, genuine interest, warmth, and a positive regard for the worth of the person. If clinicians do not demonstrate these qualities, they are unlikely to achieve the goals of the interview, no matter how these are defined.

Patient ratings of the quality of interviews have been found to be dependent on the extent to which interviewers can understand the patient's emotions and detect emotional messages that are only partially expressed, particularly as these emotions are likely to be indirect and conveyed through nonverbal behaviors. This is especially relevant in clinical interviews that focus on a client's personal difficulties. Typically, words are inadequate to accurately describe problem emotions, so interviewers must infer them from paraverbal or nonverbal expression. This is highlighted by the assumption that nonverbal aspects of communication are likely to be a more powerful method of conveying information. For example, eye contact is most likely to convey involvement; rigidity of posture might suggest client defensiveness; and hand movements often occur beyond the person's conscious intent, suggesting nervousness, intensity, or relaxation. Mehrabian (1972) has supported this perspective with his estimates that the message received is 55% dependent on facial expression, 38% by tone, and only 7% by the content of what is said.

Interviewers vary in the extent to which they take notes during the interview. Some argue that note taking during an interview might increase a client's anxiety, raise questions regarding anonymity, increase the likelihood that he or she will feel like an object under investigation, and create an unnatural atmosphere. In contrast, many interviewers counter these arguments by pointing out that a loss of rapport rarely results solely from note taking during the interview, assuming, of course, that the interviewer can still spend a sufficient amount of time attending to the client. Ongoing note taking is also likely to capture more details and result in less memory distortion than recording material after the interview has been completed. Thus, an intermediate amount of note taking during the interview is recommended. If the interview is audiotaped or videotaped, the reasons for this procedure need to be fully explained, along with the assurance of

confidentiality and the procuring of a signed agreement. Although audiotape or videotape recording is often awkward at first, usually the interviewer and client quickly forget that it is occurring.

Interview Tactics

Numerous tactics or types of statements have been proposed and studied. These include the clarification statement, verbatim playback, probing, confrontation, understanding, active listening, reflection, feedback, summary statement, random probing, self-disclosure, perception checking, use of concrete examples, and therapeutic double binds. Additional relevant topics are the importance of eye contact, self-disclosure, active listening, and touch. These areas are beyond the scope of this chapter, but the interested reader is referred to excellent discussions by Cormier and Cormier (1998), Sommers-Flanagan and Sommers-Flanagan (1999), Sattler (2002), and Zuckerman (2000). The most relevant skills for interviewing do not come so much from memorizing interviewing tactics, but develop from reviewing actual live or taped interview sessions. However, several important tactics of interviewing are described because they provide a general interviewing strategy.

Preliminaries

During the initial phase of the interview, practitioners need to ensure that they deal adequately with the following issues:

1. Organize the physical characteristics of the interview situation so that the room looks used but not untidy; lighting is optimal; seating is arranged so that the interviewer and client are neither too close nor too far and so that eye level is approximately equal.
2. Introduce yourself and indicate how you prefer to be addressed (Doctor, first name, etc.) and clarify how the client prefers to be addressed.
3. State the purpose of the interview, check the client's understanding of the interview, and clarify any discrepancies between these two understandings.
4. Explain how the information derived from the interview will be used.
5. Describe the confidential nature of the information, the limits of confidentiality, and special issues related to confidentiality (e.g., how the information might be obtained and used by the legal justice system). Further, explain that the client has the right not to discuss any information he or she does not wish to disclose. If the information will be sent to other persons, obtain a signed release of information.
6. Explain the role and activities you would like the client to engage in, the instruments that are likely to be used in the assessment, and the total length of time required. In some circumstances, this may be formalized into a written contract (Handelsman & Galvin, 1988).
7. Make sure that any fee arrangements have been clarified, including the hourly rate, total estimated cost, the amount the client versus a third party is likely to need to pay, and the interval between billing and the expected payment.

With the possible exception of fee arrangement (item 7), the preceding issues should be handled by a mental health practitioner rather than a secretary or receptionist. Covering these areas during the preliminary stages of the interview is likely to reduce the likelihood of miscommunications and later difficulties.

Directive versus Nondirective Interviews

The degree to which clinicians choose to be structured and directive during an interview depends on both theoretical and practical considerations. If time is limited, the interviewer needs to be direct and to the point. The interviewer will use a different approach for assessing a person who has been referred and will be returning to the referring person than for a person before conducting therapy with him or her. An ambiguous, unstructured approach probably makes an extremely anxious person even more anxious, while a direct approach may prove more effective. A passive, withdrawn client also is likely to initially require a more direct question-and-answer style. As stated previously, a less structured style often encourages deeper client self-exploration, enables clinicians to observe the client's organizational abilities, and may result in greater rapport, flexibility, and sensitivity to the client's uniqueness.

Frequently, behavioral interviews are characterized as being structured and directed toward obtaining a comprehensive description of actual behaviors and relevant cognitions, attitudes, and beliefs (see Chapter 4). This is often contrasted with the more unstructured psychodynamic approach, which investigates underlying motivations and hidden dynamics, and assesses information that may not be within the person's ordinary awareness. Typically, these approaches are perceived as competing and mutually exclusive. Haas, Hendin, and Singer (1987) point out that this either/or position is not only unnecessary but unproductive, because each style of interviewing provides different types of information that could potentially compensate for the other's weaknesses. Using both approaches might increase interview breadth and validity. This is similar to basing client descriptions on direct behavioral data (public communication), self-description, and private symbolization (Leary, 1957). Each of these levels may be useful for different purposes, and the findings from each level might be quite different from one another.

Sequence of Interview Tactics

Most authors recommend that interviewers begin with open-ended questions and, after observing the client's responses, use more direct questions to fill in gaps in their understanding (Beutler & Groth-Marnat, 2003; Othmer & Othmer, 1994; Sommers-Flanagan & Sommers-Flanagan, 1999). Although this sequence might begin with open-ended questions, it should typically lead to interviewer responses that are intermediate in their level of directiveness, such as facilitating comments, requesting clarification, and possibly confronting the client with inconsistencies.

An important advantage of open-ended questions is that they require clients to comprehend, organize, and express themselves with little outside structure. This is perhaps the only occasion in the assessment process that makes this requirement of clients, because most tests or structured interviews provide guidance in the form of specific, clear stimuli. When clients are asked open-ended questions, they will be most likely to express significant but unusual features about themselves. Verbal fluency, level of assertiveness, tone of voice, energy level, hesitations, and areas of anxiety can be noted.

Hypotheses can be generated from these observations and further open-ended or more direct questions used to test these hypotheses. In contrast to these advantages, open-ended questions can potentially provide an overabundance of detailed, vague, and tangential information.

Interviewer responses that show an intermediate level of directiveness are facilitation, clarification, empathy, and confrontation. Facilitation of comments maintains or encourages the flow of conversation. This might be accomplished verbally ("Tell me more . . . ," "Please continue . . .") or nonverbally (eye contact, nodding). These requests for clarification might be used when clients indicate, perhaps through subtle cues, that they have not fully expressed something regarding the topic of discussion. Requests for clarification can bring into the open material that was only implied. In particular, greater clarification might be achieved by requesting the client to be highly specific, such as asking him or her to provide concrete examples (a typical day or a day that best illustrates the problem behavior). Empathic statements ("It must have been difficult for you") can also facilitate client self-disclosure.

Sometimes interviewers might wish to confront, or at least comment on, inconsistencies in a client's information or behavior. Carkhuff (1969) has categorized the potential types of inconsistencies as being between what a person is versus what he or she wants to be, what he or she is saying versus what he or she is doing, and between the person's self-perception versus the interviewer's experience of the person. A confrontation might also challenge the improbable content of what he or she is reporting ("tall" stories).

The purpose of confrontations during assessment is to obtain more in-depth information about the client. In contrast, therapeutic confrontations are used to encourage client self-exploration and behavior change. If a practitioner is using the initial interview and assessment as a prelude to therapy, this distinction is less important. However, a confrontational style can produce considerable anxiety, which should be created only if sufficient opportunity exists to work through the anxiety. Usually, a client is most receptive to confrontations when they are posed hypothetically as possibilities to consider rather than as direct challenges. Confrontations also require a sufficient degree of rapport to be sustained; unless this rapport is present, confrontations probably result in client defensiveness and a deterioration of the relationship.

Finally, direct, close-ended questions can be used to fill in gaps in what the client has stated. Thus, a continual flow can be formed between client-directed or client-organized responses and clinician-directed responses. This sequence, beginning with open-ended questions, then moving to intermediately structured responses (facilitation, clarification, confrontation), and finally ending in directive questions, should not be rigid but should vary throughout the interview.

Comprehensiveness

The basic focus of an assessment interview should be to define the problem behavior (nature of the problem, severity, related affected areas) and its causes (conditions that worsen or alleviate it, origins, antecedents, consequences). Interviewers might wish to use a checklist, such as the one in Table 3.1, to ensure they are covering most relevant areas. In using such a checklist, the interviewer might begin with a general question, such as "How were you referred here?" or "What are some areas that concern you?" Observations and notes can then be made about the way the client organizes his or her

responses, what he or she says, and the way he or she says it. The interviewer could use facilitating, clarifying, and confronting responses to obtain more information. Finally, the interviewer could review the checklist on family background to see if all relevant areas were covered sufficiently. If some areas or aspects of areas weren't covered, the interviewer might ask direct questions, such as "What was your father's occupation?" or "When did your mother and father divorce?" The interviewer could then begin the same sequence for personal history related to infancy, middle childhood, and so on. Table 3.1 is not comprehensive, but is intended as a general guide for most interview situations. If practitioners generally evaluate specific client types (child abuse, suicide, brain-impaired), this checklist may need additional guidelines and/or be used as an adjunct to commercially available structured interviews, such as the Personality Disorder Examination (Loranger, 1988), Neuropsychological Status Examination (Schinka, 1983), or Lawrence Psychological-Forensic Examination (Lawrence, 1984).

Avoidance of "Why" Questions

It is best to avoid "why" questions because they are likely to increase client defensiveness. A "why" question typically sounds accusatory or critical and thus forces the client to account for his or her behavior. In addition, clients are likely to become intellectual in this situation, thereby separating themselves from their emotions. An alternative approach is to preface the question with either "What is your understanding of . . ." or "How did it occur that . . ." rather than "why?" These options are more likely to result in a description rather than a justification and to keep clients more centered on their emotions.

Nonverbal Behaviors

Interviewers should also be aware of their own as well as their clients' nonverbal behaviors. In particular, interviewers might express their interest by maintaining eye contact, being facially responsive, and attending verbally and nonverbally, such as through occasionally leaning forward.

Concluding the Interview

Any interview is bound by time constraints. An interviewer might help to ensure observance of these constraints by alerting the client when only 5 or 10 minutes remain until the arranged completion of the interview. This allows the client or interviewer to obtain final relevant information. There should also be an opportunity for the client to ask any questions or provide comments. At the end of an interview or assessment session, the interviewer should summarize the main themes of the interview and, if appropriate, make any recommendations.

MENTAL STATUS EXAMINATION

The mental status exam was originally modeled after the physical medical exam; just as the physical medical exam is designed to review the major organ systems, the mental status exam reviews the major systems of psychiatric functioning (appearance, cognitive function, insight, etc.). Since its introduction into American psychiatry by Adolf

Meyer in 1902, it has become the mainstay of patient evaluation in most psychiatric settings. Most psychiatrists consider it as essential to their practice as the physical examination is in general medicine (Rodenhauser & Fornal, 1991).

A mental status examination can be used as part of a formal psychological assessment for a variety of reasons. A brief mental status examination might be appropriate before assessment to determine the appropriateness of more formal psychological testing. If, for example, a patient was unable to determine where he or she was and had significant memory impairments, testing with most instruments might be too difficult and could thereby result in needless distress. Such a screening might also be used to determine basic case management issues such as hospitalization or placing the patient under close observation. A mental status examination can also be used as part of an assessment using formal psychological tests. The "raw" data from the exam can be selectively integrated with general background information to present a coherent portrait of the person and assist in diagnosis.

Despite its popularity among psychiatrists, this form of interviewing is not typically used by psychologists, partly because many areas reviewed by the mental status exam are already covered during the assessment interview and through the interpretation of psychological test results. Many psychological tests cover these areas in a more precise, in-depth, objective, and validated manner with scores being compared to appropriate norms. A client's appearance, affect, and mood are usually noted by attending to behavioral observations. A review of the history and nature of the problem is likely to pick up areas such as delusions, misinterpretations, and perceptual disorders (hallucinations). Likewise, interview data and psychological test results typically assess a client's fund of knowledge, attention, insight, memory, abstract reasoning, and level of social judgment. However, the mental status examination reviews all of the preceding areas in a relatively brief, systematic manner. Furthermore, there are situations, such as intakes in an acute medical or psychiatric hospital, where insufficient time is available to evaluate the client with psychological tests.

Numerous sources in the psychiatric literature provide thorough guidelines for conducting a mental status exam (Crary & Johnson, 1981; H. Kaplan & Sadock, 2001; Othmer & Othmer, 1994; Robinson, 2001; Sommers-Flanagan & Sommers-Flanagan, 1999), and R. Rogers (1995) has provided a review of the more structured mental status exams. This literature indicates that practitioners vary widely in how they conduct the mental status examination. The most unstructured versions involve merely the clinician's use of the mental status examination as a set of general guidelines. The more structured versions range from comprehensive instruments that assess both general psychopathology and cognitive impairment to those that focus primarily on cognitive impairment. For example, the comprehensive North Carolina Mental Status Examination (Ruegg, Ekstrom, Evans, & Golden, 1990) includes 36 items that are rated on a three-point scale (not present, slight or occasional, marked or repeated) to cover the important clinical dimensions of physical appearance, behavior, speech, thought processes, thought content, mood, affect, cognitive functioning, orientation, recent memory, immediate recall, and remote memory. Another similar comprehensive instrument is the Missouri Automated Mental Status Examination Checklist (Hedlund, Sletten, Evenson, Altman, & Cho, 1977), which requires the examiner to make ratings on the following nine areas of functioning: general

appearance, motor behavior, speech and thought, mood and affect, other emotional reactions, thought content, sensorium, intellect, and insight and judgment. The checklist includes 119 possible ratings, but the examiner makes ratings in only those areas he or she judges to be relevant.

Despite extensive development, the more comprehensive mental status examinations have not gained wide acceptance. In contrast, the narrower structured mental status examinations that focus more exclusively on cognitive impairment are used quite extensively. One of the most popular has been the Mini Mental Status Examination (Folstein, Folstein, & McHugh, 1975). It comprises 11 items designed to assess orientation, registration, attention, calculation, and language. It has excellent interrater and test-retest reliabilities (usually well above .80), correlates with WAIS IQs (.78 for verbal IQ), and is sensitive to global and left hemisphere deficits (but not right hemisphere impairment; R. Rogers, 1995; Tombaugh, McDowell, Kristjansson, & Hubley, 1996). Clinicians who wish to develop knowledge and skills in conducting mental status examinations are encouraged to consult the preceding sources.

The following descriptions of the typical areas covered serve as a brief introduction to this form of interviewing. The outline is organized around the categories recommended by Crary and Johnson (1981), and a checklist of relevant areas is included in Table 3.2. Interviewers can answer the different areas on the checklist either during or after a mental status examination. The tabled information can then be used to answer relevant questions relating to the referral question, to help in diagnosis, or to add to other test data. Such a checklist is important because clinicians not using similar checklists have been found to frequently omit crucial information (Ruegg et al., 1990).

General Appearance and Behavior

This area assesses material similar to that requested in the "behavioral observations" section of a psychological report (see Chapter 15). A client's clothing, posture, gestures, speech, personal care/hygiene, and any unusual physical features such as physical handicaps, tics, or grimaces are noted. Attention is given to the degree to which his or her behavior conforms to social expectations, but this is placed in the context of his or her culture and social position. Additional important areas are facial expressions, eye contact, activity level, degree of cooperation, physical attractiveness, and attentiveness. Is the client friendly, hostile, seductive, or indifferent? Do any bizarre behaviors or significant events occur during the interview? In particular, speech might be fast or slow, loud or soft, or include a number of additional unusual features. Table 3.2 includes a systematic checklist of relevant areas of behavior and appearance.

Feeling (Affect and Mood)

A client's *mood* refers to the dominant emotion expressed during the interview, whereas *affect* refers to the client's range of emotions. This is inferred from the content of the client's speech, facial expressions, and body movements. The type of affect can be judged according to variables such as its depth, intensity, duration, and appropriateness. The client might be cold or warm, distant or close, labile, and, as is characteristic of

Table 3.2 Format for mental status and history

Name _____ Observer's Name _____

			No Data	Present	Absent
APPEARANCE		1. unkempt, unclean, disheveled.....			
		2. clothing and/or grooming atypical.....			
		3. unusual physical characteristics.....			
COMMENTS RE APPEARANCE:					
BEHAVIOR	Posture	4. slumped.....			
		5. rigid, tense.....			
	Facial Expression Suggests	6. anxiety, fear, apprehension.....			
		7. depression, sadness.....			
		8. anger, hostility.....			
		9. absence of feeling, blandness.....			
		10. atypical, unusualness.....			
	General Body Movements	11. accelerated, increased speed.....			
		12. decreased, slowed.....			
		13. atypical, unusual.....			
		14. restlessness, fidgetiness.....			
	Speech	15. rapid speech.....			
		16. slowed speech.....			
		17. loud speech.....			
		18. soft speech.....			
		19. mute.....			
		20. atypical quality, slurring, stammer.....			
BEHAVIOR	Therapist-Patient Relationship	21. domineering, controlling.....			
		22. submissive, overly compliant, dependent..			
		23. provocative, hostile, challenging.....			
		24. suspicious, guarded, evasive.....			
		25. uncooperative, non-compliant.....			
COMMENTS RE BEHAVIOR:					
FEELING (AFFECT AND MOOD)		26. inappropriate to thought content.....			
		27. increased lability of affect.....			
		predominant mood is:			
		28. blunted, dull, bland.....			
		29. euphoria, elation.....			
		30. anger, hostility.....			
		31. anxiety, fear, apprehension.....			
		32. depression, sadness.....			
COMMENTS RE FEELING:					
PERCEPTION		33. illusions.....			
		34. auditory hallucinations.....			
		35. visual hallucinations.....			
		36. other types of hallucinations.....			
COMMENTS RE PERCEPTION:					
THINKING	Intellectual Functioning	37. impaired level of consciousness.....			
		38. impaired attention span, distractible....			
		39. impaired abstract thinking.....			
		40. impaired calculation ability.....			
		41. impaired intelligence.....			
	Orientation	42. disoriented to person.....			
		43. disoriented to place.....			
		44. disoriented to time.....			
	Memory	45. impaired recent memory.....			
		46. impaired remote memory.....			
	Insight	47. denies presence of psychological problems.....			
		48. blames others or circumstances for problems.....			
	Judgment	49. impaired ability to make routine decisions.....			
		50. impaired impulse control.....			
THINKING	Thought Content	51. obsessions.....			
		52. compulsions.....			
		53. phobias.....			
		54. depersonalization.....			
		55. suicidal ideation.....			
		56. homicidal ideation.....			
		57. delusions.....			
	Stream of Thought	58. associational disturbance.....			
COMMENTS RE THINKING:					
DIAGNOSIS: _____ as manifested by the following M.S.E. items					
_____ _____ _____ _____ _____ _____					
_____ _____ _____ _____ _____ _____					

Reproduced by permission of MTP Press LTD., Lancaster, England, from "Mental Status Examination" by W. G. Crary and C. W. Johnson, 1981. In Johnson, C. W., Snibbe, J. R., and Evans, L. A. (Eds.), *Basic Psychopathology: A Programmed Text,* 2nd ed. Lancaster: MIP Press, pp. 55–56.

schizophrenia, his or her affect might be blunted or flattened. The client's mood might also be euphoric, hostile, anxious, or depressed.

Perception

Different clients perceive themselves and their world in a wide variety of ways. It is especially important to note whether there are any illusions or hallucinations. The presence of auditory hallucinations are most characteristic of schizophrenics, whereas vivid visual hallucinations are more characteristic of persons with organic brain syndromes.

Thinking

Intellectual Functioning

Any assessment of higher intellectual functioning needs to be made in the context of a client's educational level, socioeconomic status, and familiarity and identification with a particular culture. If a low level of intellectual functioning is consistent with a general pattern of poor academic and occupational achievement, a diagnosis of intellectual disability might be supported. However, if a person performs poorly on tests of intellectual functioning and yet has a good history of achievement, organicity might be suspected.

Intellectual functioning typically involves reading and writing comprehension, general fund of knowledge, ability to do arithmetic, and the degree to which the client can interpret the meaning of proverbs. Throughout the assessment, clinicians typically note the degree to which the client's thoughts and expressions are articulate versus incoherent. Sometimes clinicians might combine assessments of intellectual functioning with some short, formal tests such as the Bender, with an aphasia screening test, or even with portions of the WAIS-III or WISC-III.

Orientation

The ability of clients to be oriented can vary in the degree to which they know who they are (person), where they are (place), and when current and past events have occurred or are occurring (time). Clinical observation indicates the most frequent type of disorientation is for time, whereas disorientation for place and person occurs less frequently. When disorientation does occur for place, and especially for person, the condition is relatively severe. Disorientation is most consistent with organic conditions. If a person is oriented in all three spheres, this is frequently abbreviated as "oriented X3."

Related to the orientation of clients is their *sensorium,* which refers to how intact their physiological processes are to receiving and integrating information. Sensorium might refer to hearing, smell, vision, and touch and might range from being clouded to clear. Can the client attend to and concentrate on the outside world or are these processes interrupted? The client might experience unusual smells, hear voices, or have the sense that his or her skin is tingling. Sensorium can also refer to the client's level of consciousness, which may vary from hyperarousal and excitement to drowsiness and confusion. Disorders of a client's sensorium often reflect organic conditions, but may also be consistent with psychosis.

Memory, Attention, and Concentration

Because memory retrieval or acquisition requires attention and concentration, these three functions are frequently considered together. Long-term memory is often assessed by requesting information regarding the client's general fund of information (e.g., important dates, major cities in a country, three major heads of state since 1900). Some clinicians include the Information or Digit Span subtests from the WAIS-III/WISC-III or other formal tests of a similar nature. Recall of a sentence or paragraph might be used to assess short-term memory for longer, more verbally meaningful information. In addition, clients' long-term memory might be evaluated by measuring recall of their major life events, and the accuracy of their recall can be compared with objective records of these events (e.g., year graduated from high school, date of marriage). It is often useful to record any significant distortions of selective recall in relation to life events as well as to note the client's attitudes toward his or her memory.

Short-term memory might be assessed by either requesting that clients recall recent events (most recent meal, how they got to the appointment) or by having them repeat digits forward and backward. Again, the WAIS-III/WISC-III Digit Span subtest, or at least a similar version of it, might be used. Serial sevens (counting forward by adding seven each time) can be used to assess how distractible or focused they are. Persons who are anxious and preoccupied have a difficult time with serial sevens as well as with repeating digits forward and, especially, backward.

Insight and Judgment

Clients vary in their ability to interpret the meaning and impact of their behavior on others. They also vary widely in their ability to provide for themselves, evaluate risks, and make plans. Adequate insight and judgment involves developing and testing hypotheses regarding their own behavior and the behavior of others. Clients also need to be assessed to determine why they believe they were referred for evaluation and, in a wider context, their attitudes toward their difficulties. How do they relate their past history to current difficulties, and how do they explain these difficulties? Where do they place the blame for their difficulties? Based on their insights, how effectively can they solve problems and make decisions?

Thinking

A client's speech can often be considered a reflection of his or her thoughts. The client's speech may be coherent, spontaneous, and comprehensible or may contain unusual features. It may be slow or fast, be characterized by sudden silences, or be loud or unusually soft. Is the client frank or evasive, open or defensive, assertive or passive, irritable, abusive, or sarcastic? Consideration of a person's thoughts is often divided into thought content and thought processes. Thought contents such as delusions might suggest a psychotic condition, but delusions may also be consistent with certain organic disorders, such as dementia or chronic amphetamine use. The presence of compulsions or obsessions should be followed up with an assessment of the client's degree of insight into the appropriateness of these thoughts and behaviors. Thought processes such as the presence of rapid changes in topics might reflect flighty ideas. The client might also

have difficulty producing a sufficient number of ideas, include an excessive number of irrelevant associations, or ramble aimlessly.

INTERPRETING INTERVIEW DATA

Interpreting and integrating interview data into the psychological report inevitably involve clinical judgment. Even with the use of structured interviews, the clinician still must determine which information to include or exclude. Thus, all the potential cautions associated with clinical judgment need to be considered (see Chapter 1). This is particularly important because life decisions and the success of later treatment may be based on conclusions and recommendations described in the report.

Several general principles can be used to interpret interview data. The interview is the primary instrument that clinicians use to develop tentative hypotheses regarding their clients. Thus, interview data can be evaluated by determining whether these hypotheses are supported by information outside the interview. Interview data that is supported by test scores can be given greater emphasis in the final report if it is relevant to the referral question. Even material that is highly supported throughout different phases of the interview process should not be included unless it relates directly to the purpose of the referral.

There is a continuum in handling interview information that varies according to the extent the information will be interpreted. On the one hand, the information might be merely reorganized into a chronological history of the person's life. This would emphasize repeating the information in as objective and accurate a manner as possible. This is typically done in the history section of a psychological report. On the other hand, interview data can be considered raw data to be interpreted. It is thus similar to the data from formal psychological tests. It might, therefore, be used to make inferences related to a client's personality, coping style, or mood and affect.

One method of organizing interview information is to use the information to develop a coherent narrative of the person's life. For example, describing how early family patterns resulted in emotionally sensitive areas ("scar" tissue) can be used to help explain current symptom patterns and difficulties in interpersonal relationships. A different sort of history might trace how interest in a vocation was first begun (early childhood daydreams regarding occupations) and how this progressed and developed as the person matured. Yet, another person might present difficulties related to authority figures. Specific details relating to these difficulties might emerge, such as the client's feeling like a martyr and eventually inappropriately expressing extreme anger toward the authority figure(s). A careful review of the client's history might reveal how he or she becomes involved in these recurring relationships and how he or she typically attempts to resolve them. Persons who are frequently depressed might distance themselves from others by their behavior and then be confused about why relationships seem to be difficult. Often, these themes emerge during a carefully conducted interview, yet aspects of the themes (or the entire themes themselves) are not apparent to the interviewee.

Interview data might also be organized around various domains (see further discussion in Chapter 15). A grid can be used to organize these domains. The various domains might be listed on the left side of the grid with the top of the grid listing the sources of data (of which the interview might be one of a variety of sources of information; see

Figure 15.1 in Chapter 15). Domains might include mood and affect, cognitions, level of resistance, symptom patterns, or coping style. This approach treats interview data in much the same manner as data from psychological tests.

There is no one strategy for sensitizing interviewers to the types and patterns of recurring themes they may encounter during interviews. Inevitably, clinical judgment is a significant factor. The accuracy and types of judgments depend on the theoretical perspective of the interviewer, knowledge regarding the particular difficulty the interviewer is investigating, past experience, types of questions asked, and purpose of the interview.

STRUCTURED INTERVIEWS

Standardized psychological tests and structured interviews were developed to reduce the problems associated with open-ended interviews. They both serve to structure the stimuli presented to the person and reduce the role of clinical judgment. Because structured interviews generate objective ratings on the same areas, they have the advantage of making possible comparisons between one case or population and the next. Typically, these interviews vary in their degree of structure, the relative expertise required to administer them, and the extent to which they serve as screening procedures designed for global measurement or as tools used to obtain specific diagnoses.

Before structured interviews could be developed, clear, specific criteria needed to be created relating to symptom patterns and diagnoses. This ideally helped to reduce the amount of error caused by vague guidelines for exclusion or inclusion in different categories *(criterion variance).* These criteria then needed to be incorporated into the interview format and interview questions. *Information variance* refers to the variability in amount and type of information derived from interviews with patients. In most unstructured interviews, information variance is caused by the wide differences in content and phrasing because of factors such as the theoretical orientation of the interviewer. Structured interviews correct for this by requesting the same or similar questions from each client.

The first popular system of specific criterion-based diagnosis was developed by Feighner et al. (1972) and provided clear, behaviorally-oriented descriptions of 16 psychiatric disorders based on the *DSM-II* (1968). Clinicians using the Feighner criteria were found to have an immediate and marked increase in interrater diagnostic reliability. The descriptions of and relevant research on the Feighner criteria were published in Woodruff, Goodwin, and Guze's (1974) book, *Psychiatric Diagnosis.* Several interviews, such as the Renard Diagnostic Interview (Helzer et al., 1981), incorporated the Feighner criteria. Spitzer, Endicott, and Robins (1978) further altered and elaborated the Feighner criteria to develop the Research Diagnostic Criteria. Simultaneous with the development of the Research Diagnostic Criteria, Endicott and Spitzer (1978) developed the Schedule for Affective Disorders and Schizophrenia (SADS), which was based on the new Research and Diagnostic Criteria. When new versions of the *DSM* were published (1980, 1987, 1994, 2000), revisions of previous interviews typically incorporated the most recent *DSM* criteria along with elements of the Feighner criteria and/or the Research Diagnostic Criteria.

As noted earlier, the reliability of structured interviews has been found to vary depending on the specificity or precision of the rating or diagnosis. Whereas the highest

reliabilities have been found for global assessment (presence/absence of psychopathology), much lower reliabilities have generally been found for the assessment of specific types of behaviors or syndromes. Likewise, high reliabilities have been found for overt behaviors, but reliability has been less satisfactory for more covert aspects of the person, such as obsessions, fears, and worries. Reliability also tends to be lower when clinicians are asked to attempt exact estimates regarding behavioral frequencies and for inferences of multifaceted aspects of the person derived from complex clinical judgments.

Most early studies on validity were based on item content (content validity) or degree of accuracy in distinguishing between broad areas of psychopathology (psychiatric/nonpsychiatric). More recent trends have attempted to assess the accuracy of far more specific areas. However, most validity studies have suffered from an absence of clear, commonly agreed-on criteria. Although structured interviews were attempts to improve on previous, imperfect instruments (unstructured interviews, standardized tests), the structured interviews themselves could not be compared with anything better. For example, the "procedural validity" strategy is based on comparing lay interviewers' diagnoses with diagnoses derived from trained psychiatrists. Although the psychiatrist's diagnosis may be better than the layperson's, diagnoses by trained psychiatrists still cannot be said to be an ultimate, objective, and completely accurate standard. Furthermore, there is confusion about whether actual validity is being measured (which would assume psychiatrists' diagnoses are the true, accurate ones) or merely a version of interrater reliability. At the core of this issue is the very nature of how diagnosis is defined and the degree to which it is actually helpful in treatment (see Beutler & Malik, 2002; Widiger & Clark, 2000).

Future studies need to involve aspects of what has previously been discussed as construct validity. This means looking more carefully at structured interviews in relationship to etiology, course, prognosis, and treatment utility relating to areas such as the appropriate selection of types of treatments and the likelihood of favorable responses to these treatments. Validity studies also need to look at the interaction between and implications of multiple criterion measures, including behavioral assessment, checklists, rating scales, self-report inventories, biochemical indices, and neuropathological alterations.

Since the mid-1970s, there has been a proliferation of structured interviews for a wide range of areas. Clinicians working in specific areas often select structured interviews directed toward diagnosing the disorders they are most likely to encounter. For example, some situations might benefit from using the Anxiety Disorders Interview Schedule (T. Brown, DiNardo, & Barlow, 1994) to make clear distinctions between anxiety disorders and substance abuse, and between psychosis and major affective disorders. Other contexts might be best served by the Eating Disorder Examination (EDE; Z. Cooper & Fairburn, 1987) or the Structured Interview for *DSM-IV* Dissociative Disorders (SCID-D; Steinberg, 1993). Three categories of structured interviews with representative frequently used instruments are included in Table 3.3 and have been extensively reviewed in R. Roger's (1995) *Diagnostic and Structured Interviewing: A Handbook for Psychologists.* One consideration in selecting these instruments is that, because most structured interviews are undergoing continuous revisions, the most up-to-date research should be consulted to ensure that practitioners obtain the most recently revised versions. The following pages provide an overview of the most frequently used and most extensively researched structured interviews.

Table 3.3 Frequently used structured interviews by categories

I.	Assessment of Axis I disorders
	Schedule of Affective Disorders and Schizophrenia (SADS)and Schedule of Affective Disorders and Schizophrenia for School-Age Children (K-SADS)
	Diagnostic Interview Schedule (DIS) and Diagnostic Interview for Children (DISC)
	Structured Clinical Interview for *DSM-III-R* (SCID)
	Diagnostic Interview for Children and Adolescents (DICA)
II.	Assessment of Axis II disorders
	Structured Interview for *DSM-III* Personality Disorders (SIDP)
	Personality Disorder Examination (PDE)
	Structured Clinical Interview for *DSM-III-R* Personality Disorders (SCID-II)
III.	Focused structured interviews
	Anxiety Disorders Interview Schedule (ADIS)
	Diagnostic Interview for Borderlines (DIB)
	Psychopathy Checklist (PCL)
	Structured Interview for *DSM-IV*-Dissociative Disorders (SCID-D)
	Structured Interview of Reported Symptoms (SIRS)
	Psychosocial Pain Inventory (PSPI)
	Comprehensive Drinker Profile (CDP)
	Eating Disorder Examination (EDE)
	Structured Interview of Sleep Disorders (SIS-D)
	Substance Use Disorders Diagnostic Schedule (SUDDS)

Schedule for Affective Disorders and Schizophrenia

The SADS (Endicott & Spitzer, 1978) is a clinician-administered, extensive, semistructured interview that has been the most widely used structured interview for clinical research purposes. Although it was originally designed for differential diagnosis between affective disorders and schizophrenia, it has evolved to include a much wider range of symptoms and allows the interviewer to consider many different diagnostic categories. Although a wide range of disorders is considered within the SADS, its primary strength lies in obtaining fine detail regarding different subtypes of affective disorders and schizophrenia (Summerfeldt & Antony, 2002). The interview rates clients on six gradations of impairment from which diagnoses are reached using the clear, objective categories derived from Spitzer et al.'s (1978) Research Diagnostic Criteria (RDC). The SADS is divided into adult versions for current symptoms, occurrence of lifetime symptoms, and degree of change. There is a further version for the assessment of children's difficulties (K-SADS). Two modifications for the SADS have been the inclusion of anxiety disorders (SADS-LA; Fyer, Endicott, Manuzza, & Klein, 1985, 1995) and eating disorders (EAT-SADS-L; Herzog, Keller, Sacks, Yeh, & Lavori, 1992).

Adult Version

The adult version of the SADS (Endicott & Spitzer, 1978) is designed to be administered in two different parts, the first focusing on the client's present illness and the

second on past episodes. This division roughly corresponds with the three different versions of the SADS. The first is the regular version (SADS), the second is the lifetime version (SADS-L, which is actually the second half of the SADS), and the third is the SADS-C, which measures changes in the client. The SADS-L is directed toward diagnosing the possible presence of psychiatric disturbance throughout the person's life. The SADS and SADS-L are the most extensively used. Because the questions in the SADS are directed toward current symptoms and those symptoms experienced one week before the illness, it is most appropriate for administration when the client is having current difficulties. In contrast, the SADS-L is most appropriate when there is no current illness. To make accurate ratings, interviewers are allowed to use a wide range of sources (client's family, medical records) and ask a number of different questions. Final ratings are made on a six-point Likert-type scale. Administration involves more than 200 items and takes from 1.5 to 2 hours and should be conducted only by a psychiatrist, psychologist, or psychiatric social worker. The end product is the following eight summary scales:

1. Mood and ideation.
2. Endogenous features.
3. Depressive-associated features.
4. Suicidal ideation and behavior.
5. Anxiety.
6. Manic syndrome.
7. Delusions-hallucinations.
8. Formal thought disorder.

Interrater reliabilities for the specific diagnostic categories have been found to be quite high, with the exception of the Formal Thought Disorder Scale (Endicott & Spitzer, 1978). The low reliability of this scale may have been because few of the patients in the Endicott and Spitzer sample showed clear patterns of disordered thoughts, which resulted in high variability for the ratings. Test-retest reliabilities were likewise good, ranging from .88 for Manic Disorders to .52 for Chronic and Intermittent Depressive Disorder (Spiker & Ehler, 1984). The exception was a low reliability for schizoaffective, depressed (.24), but this was probably because of the small number of patients included in this category, which resulted in limited variance. Using a different and possibly more appropriate statistical method, reliability increased to .84. Overall, the SADS has demonstrated excellent reliability, particularly for interrater and test-retest reliabilities related to current episodes of psychiatric disturbance.

Validity studies have been encouraging because expected relationships have been found between SADS scores and external measures of depression, anxiety, and psychosis. For example, M. H. Johnson, Margo, and Stern (1986) found that relevant SADS measures could effectively discriminate between patients with depression and paranoid and nonparanoid schizophrenia. In addition, the SADS depression measures effectively rated the relative severity of a patient's depression. For example, Coryell et al. (1994) found clear consistency between different levels of depression. The authors suggest that

incremental validity might be increased by having clients referred for a medical examination to screen out physical difficulties that might be resulting in central nervous system dysfunction. The authors also recommend that interviewers try to increase validity by always including the best available information (family history, structured tests, other rating schedules) before making final ratings. The SADS has been used to predict the clinical features, course, and outcome of various disorders, including major depression (Coryell et al., 1994), schizophrenia (Stompe, Ortwein-Swoboda, Strobl, & Friedman, 2000), and bipolar disorder (Vieta et al., 2000). A number of studies has also successfully used the SADS to detect family patterns of schizophrenia (Stompe et al., 2000) and obsessive compulsive disorders (Bienvenu et al., 2000).

Child Version

The SADS for School-Age Children (Kiddie-SADS-P, K-SADS-P; Ambrosini, 2000; Puig-Antich & Chambers, 1978) is a semistructured interview developed for children between ages 6 and 18. The K-SADS has come out in versions to be used in epidemiological research (K-SADS-E), to assess present and lifetime psychopathology (K-SADS-P/L), and present levels of symptomology (K-SADS-P). Although much of the K-SADS is based on research with major depressive disorders of prepubertal children, it also covers a wide range of disorders such as phobias, conduct disorders, obsessive-compulsive disorders, and separation anxiety.

The interview should be administered by a professional clinician who has been trained in the use of the K-SADS and is familiar with *DSM-III-R/DSM-IV* criteria. All versions are administered to both the parent and the child. Any discrepancies between the two sources of information are clarified before final ratings are made. Total administration time is approximately 1.5 hours per informant (3 hours total). The first phase is a 15- to 20-minute unstructured interview in which rapport is developed as well as an overview of relevant aspects of history, including the frequency and duration of presenting symptoms, their onset, and whether the parents have sought previous treatment. This is followed by structured questions regarding symptoms, which are rated on a Likert scale, with 1 representing "not at all" and 7 indicating that they are "extreme." A skip structure is built into the format so that interviewers can omit irrelevant questions. Interviewers are allowed to use their judgment regarding the wording and the type and number of questions. Finally, ratings are made regarding behavioral observations (appearance, attention, affect). Interviewers are also asked to rate the completeness and reliability of the interview and to make a global assessment of pathology (degree of symptomatology and level of impairment).

Test-retest and interrater reliability for the K-SADS has been good with a general trend for each version to have improved reliabilities. Ambrosini (2000), for example, reported that the K-SADS-P/L had test-retest reliabilities ranging from 1.00 (lifetime occurrence of major depression) to .55 (for lifetime occurrence for attention deficit disorder). However, overall reliabilities have been lower for the K-SADS (and K-SADS-III-R) than for the adult SADS, but this is to be expected given the relative changeableness and less well-developed language skills found with children (Ambrosini, Metz, Prabucki, & Lee, 1989; Chambers et al., 1985). Validity studies indicate that relevant K-SADS measures correlated highly with diagnoses for conduct disorders, schizophrenia,

and depression (Apter, Bleich, Plutchik, Mendelsohn, & Tyrano, 1988). Additional expected correlations have been found between SADS measures and ratings of adolescent mood (E. Costello, Benjamin, Angold, & Silver, 1991) and the Child Behavior Checklist (Achenbach & Edelbrock, 1983; Ambrosini, 2000). Finally, follow-up studies on adolescents diagnosed with disorders (i.e., depression) have found a continued risk for later affective difficulties (i.e., Lewinsohn, Rohde, Klein, & Seeley, 1999).

Collectively, the different versions of the SADS provide a thorough, well-organized interview with unparalleled coverage of the subtypes and gradations of the severity of mood disorders. The SADS has also been well accepted in research and clinical settings. It has strong interrater reliability and provides good ratings of symptom severity, measures associated symptoms, includes guidelines for possible malingering, and has strong evidence of convergent validity (see R. Rogers, 1995; Summerfeldt & Antony, 2002). In contrast, its weaknesses include a relatively narrow band of diagnosis compared with some of the other available instruments such as the SCID or DIS. In addition, the diagnoses are based on Research Diagnostic Criteria (RDC) rather than the more recent *DSM-III-R* or *DSM-IV* criteria. This criticism is somewhat moderated, however, by many of the RDC and *DSM-III/DSM-IV* criteria being nearly the same, especially for childhood disorders. Finally, administration and interpretation of the SADS require extensive training (usually a week) as well as a good working knowledge of differences between the SADS/RDC and *DSM-III-R/DSM-IV* criteria.

Diagnostic Interview Schedule

In contrast to the SADS, which is semistructured and requires administration by trained professionals, the DIS (Robins, Helzer, Croughan, & Ratcliff, 1981) is highly structured and was designed specifically by the National Institute of Mental Health (Division of Biometry and Epidemiology) to be administered by nonprofessional interviewers for epidemiological studies (see Helzer & Robins, 1988). It has been updated for the *DSM-III-R* (DIS-III-R; Robins et al., 1989) and the *DSM-IV* (DIS-IV; Robins, Cottler, Bucholz, & Compton, 1996). The latest version (DIS-IV) includes 19 modules with more than 30 Axis I diagnoses and one Axis II diagnosis (antisocial personality). This modular format allows for tailoring various portions of the DIS-IV to the interests of the researcher or clinician. However, clinical judgment is reduced to a minimum by using verbatim wording, specific guidelines, a clear flow from one question to the next, and simple yes-no answers. Thus, the DIS is far more economical to administer than the SADS. Total administration time is 60 to 90 minutes. Studies have generally indicated that results are comparable between trained clinicians and nonprofessional interviewers (Helzer, Spitznagel, & McEvoy, 1987).

Adult Version

The original version of the DIS was derived from the format of the earlier Renard Diagnostic Interview. However, diagnosis for the DIS-IV is based exclusively on *DSM-IV* criteria. Initially, questions are directed toward obtaining information regarding the client's life, and information is also requested regarding more current symptoms based on the past two weeks, past month, past six months, and past year. Specific probe questions distinguish whether a symptom is clinically significant. A total of 470 potential

clinical ratings are made and organized around 24 major categories. Administration time is approximately 60 to 90 minutes.

Computerized administration and scoring programs are available that can generate *DSM-IV*-based diagnoses. However, computer-based diagnoses on early versions of the DIS were found to generate an average of 5.5 possible diagnoses compared with an average of 2.6 for nonstructured interviews (Wyndowe, 1987). Patient acceptance for the computer administration has been found to be high, although the average administration time is somewhat longer than the clinician-interviewed version.

Studies of the reliability and validity of the DIS have been both variable and controversial. Although much of this research was done on pre-DIS-IV versions, the similarity of format and content between the DIS and DIS-IV suggests that much of this earlier research is pertinent. The comparability of diagnosis by professionals and nonprofessionals using the DIS has generally been supported. This suggests that nonprofessionals can effectively use it to help gather data for large epidemiological studies. For example, Robins et al. (1981) found diagnostic agreement between psychiatrists and nonprofessional interviewers to be .69. The sensitivity (percent interviewees correctly identified) of the DIS varied according to type of diagnosis, but had a mean of 75% with a mean specificity (percent noncases correctly identified) of 94%. More recent studies have similarly concluded that the specificity is stronger than its sensitivity (Eaton, Neufeld, Chen, & Cai, 2000; J. Murphy, Monson, Laird, Sobol, & Leighton, 2000). However, data on sensitivity and specificity were based on using psychiatrists' diagnoses as the true index of diagnostic accuracy. The difficulties in considering psychiatrists' ratings as the truly accurate or "gold standard" criterion for validity have already been noted; therefore, it is probably best to consider the preceding data on sensitivity and specificity as forms of interrater reliability rather than concurrent validity. In contrast to this study, Vandiver and Sheer (1991) found somewhat modest median test-retest reliabilities ranging between .37 and .46.

Although many of the DIS ratings between professional and lay interviewers were equivalent, Helzer et al. (1985) found that, when compared with psychiatrists, nonprofessional interviewers tended to overdiagnose major depression. In contrast to Helzer et al. (1987), Folstein et al. (1985) did not find a sufficiently high rate of agreement between diagnoses by a panel of psychiatrists and diagnoses by the DIS to warrant its use in epidemiological studies. Specifically, it was found that the DIS generated more cases of depression and schizophrenia and fewer cases of alcoholism and antisocial personality (Cooney, Kadden, & Litt, 1990; Folstein et al., 1985). Eaton et al. (2000) has noted that false-negative diagnoses for many cases could be attributed mainly to failure by patients to report symptoms based on life crises or medical conditions. In contrast, the DIS has been found to be comparable with other commonly used psychiatric rating devices such as the Psychiatric Diagnostic Interview (Folstein et al., 1985; R. Weller et al., 1985). However, both diagnostic strategies may contain inaccuracies, and it is difficult to tell in which areas these inaccuracies occurred (R. Weller et al., 1985). The DIS has had the greatest difficulty accurately diagnosing borderline conditions and patients in remission, but this is to be expected because these are the most problematic diagnoses for many other assessment strategies (Robins & Helzer, 1994). In contrast, Swartz et al. (1989) were able to find quite respectable sensitivities (85.7%) and specificities (86.2%) for borderline conditions using a DIS borderline index.

Child Version

The Diagnostic Interview Schedule for Children (DISC; Costello, Edelbrock, Duncan, & Kalas, 1984; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) is similar to the adult version in that it is highly structured and designed for nonprofessional interviewers. It differs in that it is designed to be given as both a child interview (DISC-C) and parent interview (DISC-P). There have also been versions designed for teachers (Teacher DISC), screening (DISC Predictive Scales), young adults (Young Adult DISC), and administrations that can be given by computer or audiotape (Lucas et al., 2001; Shaffer et al., 2000). Ratings are coded as 0 (not true), 1 (somewhat true), or 2 (very often true). *DSM-IV* diagnoses are generated based on the combined ratings for the child and parent interviews. Some of the more problematic diagnoses (autism, pervasive developmental disorder, pica) are based on an interview with the parent only. The entire interview takes an average of 70 minutes per informant and 90 to 120 minutes per patient, but an explicit skip structure can enable some interviews to be somewhat shorter. The most recent modification of the DISC (DISC-IV; Robins et al., 1996; Shaffer et al., 2000) was designed to be compatible with *DSM-IV* and *ICD-10* criteria. The DISC-IV comprises six modules, each of which comprises the major diagnostic clusters (Anxiety, Mood, Disruptive, Substance Use, Schizophrenia, Miscellaneous).

DISC test-retest reliability (one-year interval) for *DSM-IV* diagnoses in a clinical sample was good to adequate with parent ratings having higher reliabilities (.54 to .79) than child interviews (.25 to .92; Shaffer et al., 2000). However, test-retest reliabilities for a community sample were generally quite poor for child interviews (.27 to .64) but adequate for parent interviews (.45 to .68; Shaffer et al., 2000). Children's reliability increased with age, which is expected considering their increase in intellectual abilities, greater memory, and improved language comprehension and expression. In contrast, reliabilities based on ratings from interviews with the parents decreased with the child's age, probably because the parents have progressively less contact with their child.

Research on the validity of the DISC has found that discriminations between psychiatric and pediatric groups were good for children with severe diagnoses and severe symptoms but not for children with mild-to-moderate difficulties (Shaffer et al., 2000). Discriminations based on interviews with parents were generally more accurate than those based on children (E. Costello, Edelbrock, & Costello, 1985). Accuracy was also higher for externalizing than internalizing disorders (Friman et al., 2000). In addition, comparisons between psychiatric and pediatric referrals indicated that psychiatric referrals had more symptom scores and more psychiatric diagnoses than pediatric referrals (E. Costello et al., 1985). The DISC has also been found to identify risk factors for substance abuse (Greenbaum, Prange, Friedman, & Silver, 1991) and to predict behaviors related to conduct and oppositional disorders (Friman et al., 2000). Ratings between DISC and clinician-based diagnosis were moderate to good (.29 to .74 for parent and .27 to .79 for child; Shaffer et al., 2000) in research settings and followed strict diagnostic guidelines. However, there was very poor agreement between DISC and clinician-based diagnosis when the clinicians performed diagnosis in everyday clinical settings (A. L. Jensen & Weisz, 2002). This may reflect not so much a weakness of the DISC itself, but more that there are considerable differences between how diagnosis is achieved in research as opposed to practice contexts. In summary, the DISC has strengths in that it has

good reliability and validity among clinical samples involving parent interviews, especially when the problems are related to externalizing disorders. However, the DISC is more problematic when ratings are based on child interviews, particularly among community samples and for internalizing disorders.

Diagnostic Interview for Children and Adolescents

The Renard Diagnostic Interview (Helzer et al., 1981) inspired both the DIS and the Diagnostic Interview for Children and Adolescents (DICA; Herjanic & Campbell, 1977; Herjanic & Reich, 1982). It has been through several revisions, which have incorporated the different editions of the *DSM* and elements of the DIS (W. Reich, 2000). Similar to the DIS, the DICA has been designed for administration by lay interviewers. The most recent version was published in 1997 and is available in child, adolescent, and parent versions (W. Reich, 2000). The DICA can be administered to children between ages 6 and 17 years. The format is semistructured and primarily organized around different themes, such as behavior at home, behavior at school, and interpersonal relationships with peers. Additional content areas are substance abuse and the presence of syndromes such as anxiety disorders, mania, and affective disorders. Elaborate instructions are given for skipping irrelevant items, and total administration time is between one to two hours. The administration begins with an interview of both the parent and child, which is designed to establish baseline behaviors and to obtain relevant chronological information. The parent is then questioned about the child to determine the possible appropriateness of common *DSM-IV* diagnostic categories. The final step is to administer a "Parent Questionnaire," which requests additional medical and developmental history and addresses possible diagnoses that have not been covered by previous questioning.

Reliability of the DICA has been quite variable. Test-retest reliability has been quite good, mostly ranging between .76 and .90 (Bartlett, Schleifer, Johnson, & Keller, 1991; Earls, Reich, Jung, & Cloninger, 1988). However, test-retest reliability for child (6 to 12) ADHD was low (.32) and oppositional disorder was low to adequate (.46; W. Reich, 2000). Reliability has been found to be lowest for questions that were complex, related to time, and for children with the highest level of functional impairment. In contrast, questions with the highest reliability were related to frequency and to externalizing symptoms (Perez, Ascaso, Massons, & Chaparro, 1998). Most cross-informant (parent-child) agreement related to specific symptoms has been disappointingly low (.19 to .54; Herjanic & Reich, 1982). The highest level of agreement was for the oldest children and the lowest for younger groups (W. Reich, 2000). Whereas mothers reported more behavioral symptoms, children were more likely to report subjective complaints.

Validity studies on the DICA indicate that it can accurately make the somewhat gross distinction between middle- to older-aged children who were referred to a general psychiatric clinic from those referred to a pediatric clinic (Herjanic & Campbell, 1977). However, there was considerable overlap for children between ages six and eight, thus suggesting that a greater possibility of misdiagnosis exists for children in this age range. The DICA was found to be most effective for assessing relationship problems, less effective for academic difficulties, and least effective for assessing school problems, somatic complaints, and neurotic symptoms (Herjanic & Campbell, 1977). In addition,

adolescents diagnosed with depression on the DICA also had corresponding elevations on the Beck Depression Inventory (Martin, Churchard, Kutcher, & Korenblum, 1991). W. Reich (2000) reported that as the genetic similarity of persons diagnosed with bipolar disorder decreased, their level of psychopathology on the DISC correspondingly decreased. In summary, the psychometric properties of the DICA have been variable with more studies needed to substantiate its validity, particularly concurrent validity (R. Rogers, 1995).

Structured Clinical Interview for the *DSM-IV*

The SCID (First, Spitzer, Gibbon, & Williams, 1996, 1997; Spitzer et al., 1987) is a clinician-administered, comprehensive broad-spectrum instrument that adheres closely to the *DSM-IV* decision trees for psychiatric diagnosis. A certain degree of flexibility is built in so that administration can be tailored to different populations and contexts. Thus, slightly different forms are used for psychiatric patients (SCID-In/Patient), outpatients (SCID-Out/Patients), and nonpatients (SCID-Non/Patients). Criticisms that the early version of the SCID had sacrificed clinical information so that it would be more user-friendly for clinicians resulted in a clear, easy-to-use version for clinical contexts (the SCID-Clinical Version; First et al., 1997) and a longer, more in-depth version for research (SCID-I; First, Spitzer, et al., 1996). Whereas these versions of the SCID are directed toward Axis I diagnoses, a separate version has been developed for the diagnosis of Axis II disorders (SCID-II; Spitzer, Williams, Gibbon, & First, 1990). A further variation, the SCID-D (Steinberg, 1993), has been developed using *DSM-IV* criteria for the assessment of dissociative disorders. The SCID and its variations include several open-ended questions as well as a skip structure, which enables the interviewer to branch into new areas dependent on the client's previous responses. Because clinical judgment is essential throughout the interview, it should be administered only by trained professionals. To increase incremental validity, the authors encourage the inclusion of relevant additional data in making final diagnostic decisions.

The SCID, along with its variations, is the most comprehensive structured interview available. As a result, administration time can be considerable even with the inbuilt screening questions and skip structure. Many individual clinicians and treatment sites deal with this by primarily administering the modules they are most concerned with. For example, a treatment center specializing in substance abuse might administer the module for Psychoactive Substance Use Disorders along with the SCID-II when the comorbidity of personality disorders is suspected. Administration time might also be reduced by administering the computerized mini-SCID (First, Gibbon, Williams, & Spitzer, 1996) that has been designed to screen for possible Axis I disorders. In addition, a computerized SCID-II (AutoSCID-II; First, Gibbon, et al., 1996) that can also potentially reduce clinician time is available. Although it can be administered by telephone, this procedure is discouraged given the poor agreement between telephone and face-to-face diagnoses (Cacciola, Alterman, Rutherford, McKay, & May, 1999).

The SCID and its variations have not been subjected to the level of reliability and validity studies as the SADS or DIS. This might be partially because of the considerable breadth of coverage encompassed by the SCID, making it a daunting task to cover all areas. The reliability studies that have been performed have resulted in overall

moderate, but quite variable, test-retest and interrater reliabilities. For example, interrater agreement using the SCID-II for common diagnostic categories ranges between .40 and .86 with a mean of .59 (First, Spitzer, Gibbon, & Williams, 1995). Riskind, Beck, Berchick, Brown, and Steer (1987) found that several difficult-to-distinguish diagnostic categories had relatively good levels of interrater agreement. These included generalized anxiety disorders (.79, 86% agreement), depressive disorders (.72, 82% agreement; Riskind et al., 1987), panic disorders ($k = .86$), and major depression ($k = .81$; J. Reich & Noyes, 1987). Test-retest reliabilities over a two-week interval for psychiatric patients was fair to good (overall weighted kappas = .61) but poor for nonpatients (overall weighted kappas = .37; J. B. Williams et al., 1992).

For the most part, validity studies of the SCID have assumed that *DSM-IV* diagnoses are the benchmark for making comparisons of diagnostic accuracy. Thus, "procedural validity" has often been assumed since the SCID has closely paralleled the diagnostic criteria derived from the *DSM-IV* (R. Rogers, 1995). A representative validity study found good agreement ($k = .83$) between interviewer ratings and cross ratings of interviewer videotapes by two senior psychiatrists (Maziade et al., 1992). Other studies have found considerable diagnostic overlap within Axis I disorders and between Axis I and Axis II disorders (Alnacs & Torgerson, 1989; Brawman-Mintzer et al., 1993). However, evaluating the meaning of this overlap is difficult because the extent to which it is caused by instrument error versus true comorbidity (i.e., the frequent occurrence of anxiety and depression) is difficult to determine. In contrast to these mostly favorable studies, a number of studies have found generally poor agreement between SCID and clinician-based diagnosis (Shear et al., 2000; Steiner, Tebes, Sledge, & Walker, 1995). In summary, the strength of the SCID is its impressive breadth of coverage, use of modules targeted toward specific areas, and close parallel with the *DSM-IV*. Its weaknesses are its wide variation in reliability and its need for further validity studies, particularly relating it to other diagnostic measures.

RECOMMENDED READING

Cormier, W. H., & Cormier, L. S. (1998). *Interviewing strategies for helpers: Fundamental skills and cognitive behavioral interventions* (4th ed.). Monterey, CA: Brooks-Cole.

Othmer, E., & Othmer, S. C. (1994). *The clinical interview using DSM-IV. Vol. 1: Fundamentals.* Washington, DC: American Psychiatric Press.

Rogers, R. (1995). *Diagnostic and structured interviewing: A handbook for psychologists.* Odessa, FL: Psychological Assessment Resources.

Sommers-Flanagan, R., & Sommers-Flanagan, J. (1999). *Clinical interviewing* (2nd ed.). New York: Wiley.

Summerdfeldt, L. J., & Antony, M. (2002). Structured and semistructured diagnostic interviews. In M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders.* New York: Guilford Press.

Chapter 4

BEHAVIORAL ASSESSMENT

Behavioral assessment is one of a variety of assessment traditions such as projective testing, neuropsychological assessment, and objective testing. Behavioral assessment distinguishes itself by being a set of specific techniques as well as a way of thinking about behavior disorders and how these disorders can be changed. One of its core assumptions is that behavior can be most effectively understood by focusing on preceding events and resulting consequences. Out of this core assumption has come a surprisingly diverse number of assessment methods, including behavioral interviewing, several strategies of behavioral observation, measurement of relevant cognitions, psychophysiological assessment, and a variety of self-report inventories.

Behavioral assessment can be most clearly defined by contrasting it with traditional assessment. One of the most important comparisons is the emphasis that behavioral assessment places on situational determinants of behavior. This emphasis means that behavioral assessment is concerned with a full understanding of the relevant antecedents and consequences of behavior. In contrast, traditional assessment is often perceived as more likely to view behavior as the result of enduring, underlying traits. It is this underlying difference in conceptions of causation that explains most of the other contrasts between the two traditions. An extension of this conceptual difference is that behavioral assessment goes beyond the attempt to understand the contextual or situational features of behavior and, more importantly, concerns itself with ways to change these behaviors. There is a close connection between assessment itself and its implications for treatment. Thus, behavioral assessment is more direct, utilitarian, and functional.

The perceived limitations of traditional assessment were a major factor in stimulating the development of behavioral assessment. Specifically, traditional assessment was considered to focus too extensively on abstract, unobservable phenomena that were distant from the actual world of the client. In addition, behaviorists felt that traditional clinical psychology had stagnated because its interventions were not sufficiently powerful and too much emphasis was placed on verbal therapy. The concepts of traditional assessment seemed to exist in an abstract world divorced from the immediate realities and requirements of behavior change. The result of many traditional procedures seemed to be a large quantity of information that had little direct relevance to treatment intervention and outcome. However, this is a stereotyped, somewhat polarized view of traditional (and behavioral) assessment in that there has been considerable and increasing emphasis on the treatment implications and situational context of information derived from traditional methods of assessment. This stereotyped view is meant to highlight differences between the two strategies rather than to capture the complexities and similarities between them.

A further contrast between behavioral and traditional assessment is that behavioral assessment is concerned with clearly observable aspects in the way a person interacts with his or her environment. A typical behavioral assessment might include specific *measures of behavior* (overt and covert), *antecedents* (internal and external), *conditions surrounding behaviors,* and *consequences.* This knowledge can then be used to specify methods for changing relevant behaviors. Although some behavioral assessors might take selected personality traits into account, these traits would be considered relevant only if they had direct implications for therapy. For example, certain personality styles interact with the extent and type of depressive cognitions (Alloy et al., 1999), and the existence of a personality disorder typically predicts therapeutic outcome (see Nelson-Gray & Farmer, 1999). This focus on the person and his or her unique situation is quite different from psychodynamic, biochemical, genetic, or normative trait models.

The behavioral approach stresses that different behavior disorders are typically expressed in a variety of modes. These might include overt behaviors, cognitions, changes in physiological states, and patterns of verbal expressions. This implies that different assessment strategies should be used for each of these modes (S. Haynes & O'Brien, 2000). An inference based on one mode does not necessarily generalize to another. For example, anxiety for one person may be caused and maintained primarily by the person's cognitions and only minimally by poor social skills. Another person might have few cognitions relating to anxiety but be anxious largely because of inadequate social skills. The person with inadequate social skills might be most effectively treated through social skills training and only minimally helped through approaches that alter irrational thoughts (see Breitholtz, Johansson, & Ost, 1999). It should also be noted that altering a person's behavior in one mode is likely to affect other modes, and these effects might have to be considered.

Whereas the preceding information presents a relatively rigid and stereotyped distinction between traditional and behavioral assessment, most practicing clinicians, including those who identify themselves as behavior therapists, typically combine and adopt techniques from both traditions (Fernandez-Ballesteros & Staats, 1992; S. Haynes & O'Brien, 2000). This is consistent with the finding that between 50% and 80% of clinicians who describe themselves as being behaviorally oriented reported using structured personality tests such as the MMPI (Guevremont & Spiegler, 1990; Watkins, Campbell, & McGregor, 1990). Watkins et al. even found that about 50% used projective tests and the Rorschach was used by a full 32%. Thus, behavioral assessment has become increasingly eclectic and now is usually perceived as part of mainstream assessment rather than as a new and contrasting alternative. Traditional and behavioral approaches have now come to resemble each other in many areas. In particular, behavioral assessment has gone through both a turning inward and a turning outward toward traditional psychometric approaches. The turning inward is most apparent in that internal behavioral repertoires and aspects of cognition are seen as essential for a complete understanding of the person (Alloy et al., 1999; Glass & Merluzzi, 2000; Linscott & DiGiuseppe, 1998; Lodge, Tripp, & Harte, 2000). Specific cognitive techniques include having the person think aloud as he or she is involved in a specific situation, sampling thoughts when a beeper goes off, and using a wide variety of self-statement inventories. Second, behavioral assessment has turned outward in that it has become increasingly concerned with traditional psychometric considerations. This has included evaluating the reliability and

validity of behavioral observations, self-report inventories, and diagnoses (Cone, 1998; Nelson-Gray & Farmer, 1999).

The assumptions and perspectives of behavioral assessment have resulted in an extremely diverse number of approaches and an even wider variety of specific techniques. These approaches and their corresponding techniques can be organized into the areas of behavioral interviewing, behavioral observation, cognitive behavioral assessment, psychophysiological assessment, and self-report inventories. Each of these areas was developed within a wider historical context extending over several decades.

HISTORY AND DEVELOPMENT

Treatment based on behavioral principles has a long history, dating back to the days of Little Albert and his fear of white, furry objects (M. Jones, 1924). However, extensive, well-defined behavioral assessment strategies that were consistent with behavioral therapy were relatively slow to develop. The earliest formal use of behavioral assessment occurred in industrial and organizational settings (Hartshorne & May, 1928; Office of Strategic Services Staff, 1948), but behavioral assessment did not become popular in the clinical context until the mid- to late 1960s. This was probably because of the powerful influence of psychodynamic approaches among clinicians who were taught to "look beneath the surface" to understand the "true" causes of behavior. Perhaps in part as a reaction to this indirect and inferential approach to understanding the person, the earliest forms of behavioral assessment focused almost exclusively on observable behaviors. Although organismic variables such as cognitions, feelings, and psychophysiological responses were acknowledged, they were not considered important influences on behavior and, as a result, were not stressed in assessment and treatment. Instead, behavioral assessment was consistent with the then-dominant operant conditioning paradigm in that it focused on identifying discrete behavioral responses, target behaviors, and reinforcers that could change specific behaviors. Measurement of these areas typically quantified the frequency, rate, and duration of relevant behaviors (Ullman & Krasner, 1965). The result was numerous, highly innovative assessments of overt behaviors. Typically, interventions involved single cases, which was consistent with their idiographic approach.

Early definitions of behavioral assessment were created partially by making contrasts with traditional psychodynamic approaches. Each had different aims (identification of problem behaviors vs. classification), assumptions (behavior is caused by situations vs. enduring traits), and applications (direct observation vs. indirect inferences). In particular, Mischel (1968) attacked the very nature of traits by arguing that they were fictions based on distortions of language (a preponderance of static descriptions), the result of consistency of roles and situations (not inner traits), perceptual bias based on needs for predictability, and the rarity of disconfirmation when traits are (incorrectly) inferred. This attack fueled a lengthy controversy, which was relevant to behavioral assessment in that Mischel's perspective was used to argue for a focus on situational determinants of behavior. Proponents of behavioral assessment (along with psychiatry itself) were also dissatisfied with traditional *DSM-II* diagnosis, which had poor reliability and validity and did not seem to relate to the real world of the client or have direct treatment utility.

During the 1970s, there was a much greater emphasis on a wider approach. The typical single case study format gave way to assessment within a much larger context such as schools, businesses, families, and differing sociocultural frameworks. This assessment approach was based partially on the observation that these larger contexts could have considerable influence on the person, so that effective individual change often required change in these wider contexts. A refocusing on larger contexts was also motivated by challenges to the strict operant paradigm in that, while effective in controlled situations (hospital ward, Skinner box, prison), it had questionable social validity and doubtful long-term clinical impact (Goldfried, 1983; Milne, 1984). Assessment was also widened by arguments to focus on the wider aspects of the person, which meant not only behavior, but also feelings, sensations, internal imagery, cognitions, interpersonal relations, and psychophysiological functioning (Lazarus, 1973). This emphasis on a multimodal or multifaceted approach forced the mainstream of behavioral assessment to accept a number of indirect measures such as self-reports, ratings by significant others, and cognitions (Cone, 1977, 1978). Relevant publications were the first editions of *Behavioral Assessment: A Practical Handbook* (Hersen & Bellack, 1976), *Handbook of Behavioral Assessment* (Ciminero, Calhoun, & Adams, 1977), and the journals *Behavioral Assessment* and the *Journal of Behavioral Assessment,* both of which began in 1979.

The 1980s and 1990s have seen a proliferation of publications in the field of behavioral assessment, a dramatic reevaluation of some of its most basic assumptions, and the incorporation of influences from other traditions and disciplines. In particular, psychiatry had similar difficulties with the *DSM-II* as behavioral assessment and began to develop strategies quite similar to those of behavioral assessment. The Problem Oriented Record (Weed, 1968) was introduced into many general hospital and psychiatric settings to improve diagnostic and treatment practices by providing behavior-specific databases, problem lists, treatment plans, and follow-up data. It thereby more effectively tied in the relationship between assessment and treatment, and more clearly delineated diagnostic issues. Perhaps of greater importance, *DSM-III-R,* and *DSM-IV* were similar to the efforts of behavioral assessment in that each diagnostic category was developed using behavior-specific descriptions. Numerous publications have worked to integrate behavioral assessment with traditional psychiatric diagnosis (First et al., 1992; Follette & Hayes, 1992; Hersen, 1988; Hersen & Bellack, 1988) in areas such as depression (R. Nelson & Maser, 1988), the diagnosis of childhood disorders (Kazdin, 1988), personality disorders (Nelson-Gray & Farmer, 1999), and understanding different models of causation (S. Haynes & O'Brien, 1988). The perspectives of psychiatry and behavioral assessment have been further linked by the *Journal of Behavior Therapy and Experimental Psychiatry.*

The development and expansion of behavioral medicine has also drawn extensively on behavioral assessment strategies in the evaluation of headaches, coronary heart disease, Reynaud's disease, asthma, chronic pain, sleep disturbances, and eating disorders (Williamson, Veron-Guidry, & Kiper, 1998). More recently, behavioral assessment strategies have begun to focus on unstable, transitional behaviors in part motivated by new conceptual developments based on chaos theory (S. Haynes, 1995). Thus, not only has behavioral assessment increasingly accepted the contributions of other disciplines and alternative models of conceptualizing behavior, but many of the most honored behavioral techniques have been challenged (Goldfried, 1983). For example, clinical judgment

in the context of structured interviews has been accepted, diagnostic classification is now considered potentially useful, reliance solely on behavioral observations is perceived in some contexts as inappropriate, and indirect measurement is often seen as essential. In addition, more inferential techniques such as measuring underlying cognitive structures (schemas) that organize more specific thoughts and behaviors, have now become a frequent part of behavioral assessment (Linscott & DiGiuseppe, 1998). This is contrasted by a dramatic decrease in the early, time-honored focus on measuring observable frequencies of target behaviors (Glass & Merluzzi, 2000; Guevremont & Spiegler, 1990).

In essence, the 1980s and 1990s witnessed a significant reappraisal and expansion of what is involved in behavioral assessment. Birchler (1989) summarizes his review by noting, "Behavioral assessment as we may have known it in the recent past is in a rapidly changing process of (choose one): disarray, revision, broad expansion, advancement, confusion, and/or extinction" (p. 385). There has certainly been a significant blurring and cross-fertilization between behavioral assessment and other forms of assessment (S. Haynes & O'Brien, 2000). This is in part reflected in the fact that the *Behavioral Assessment* journal and the *Journal of Behavioral Assessment* have now changed their names and content to include wider aspects of psychopathology and more traditional assessment tools (i.e., MMPI-2, MCMI-III). This integration and overlap seem to be partially based on the belief that predicting behavior and optimizing treatment outcomes is probably most effective when the strengths of both traditions are used—that each contains complementary types of information.

ISSUES RELATED TO RELIABILITY AND VALIDITY

Traditional psychometric considerations for behavioral assessment are difficult to summarize because of the wide diversity of techniques and the differences in assumptions regarding the focus, nature, and causes of behavior. Whereas traditional assessment stresses the relative stability of various characteristics, behavioral assessment assumes variability based largely on environmental factors. A finding such as low test-retest reliability is more likely to be interpreted in the behavioral context because of true variance resulting from environmental conditions rather than error in the data collection procedure. Furthermore, behavioral assessment stresses the importance of individually tailored approaches emphasizing the client's idiosyncrasies. In this context, normative comparisons are frequently seen as both irrelevant and inappropriate. Despite these issues, many from within the area of behavioral assessment have successfully argued for evaluating behavioral assessment techniques with traditional psychometric approaches (T. Anderson, Cancelli, & Kratochwill, 1984; Cone, 1998). For example, interobserver agreement for behavioral observations is essential before the data gathered from this approach can be trusted. This is typically determined by calculating the percentage of interrater agreement. Likewise, data derived from self-reports in areas such as assertiveness and fear need to demonstrate that the findings can be generalized to other situations such as role plays, simulations, and, especially, daily life.

The earliest forms of behavioral assessment relied primarily on behavioral observation and assumed that the direct observation of specific behaviors was sufficiently

clear, reliable, and accurate. The emphasis was primarily on determining a functional analysis between behavior and its antecedents and consequences. In an activity such as pressing a bar for reinforcement, the behavior could be easily recorded by an electronic detector, and, therefore, the reliability of the measure could be considered to be quite high. However, with behaviors that are more difficult to define, the reliability of measurement, especially when based on behavioral observation, cannot be assumed. For example, fingernail-biting might be defined merely by the person touching his or her face, or it may involve touching the mouth, actually chewing the nail, or removing part of the nail or perhaps the entire nail. The issue of precise definition and accurate measurement of the behavior becomes even more problematic when dealing with internal cognitions, in which the clinician is completely dependent on self-reports rather than on direct observation.

The level of reliability across different observational strategies has varied. In general, material derived from behavioral observation during behavioral assessment can be influenced by observer expectations in similar ways, as has been found by experimental research (H. Cooper & Rosenthal, 1980; Orne, 1962; R. Rosenthal, 1966). Consistent with this is that interrater agreement has been quite variable for areas such as overt difficulties and underlying mechanisms (Persons, Mooney, & Padesky, 1995). In situations such as natural observation in which observer bias, outside factors (such as interference from nontarget persons), and a lack of clear definitions are likely to create variability in observer ratings, reliability can be expected to be relatively low. Further sources of observer error include halo effects, primacy effects, failure to score a behavior that has occurred, rating toward the center of the scale, and leniency or generosity of scoring. When bias is reduced by using highly structured procedures, reliability increases. Thus, a procedure such as systematic sampling in which clear strategies are used to determine when and how the behavior will be measured has generally been more reliable and accurate than naturalistic observation (Cunningham & Thorp, 1981). Although reliability has been found to increase in controlled situations where the observers know that they, themselves, are being evaluated for accuracy (Romanczyk, Kent, Diament, & O'Leary, 1973), this outside monitoring of observers rarely occurs in clinical situations. Thus, it cannot be assumed that the reliability found in clinical situations is as high as for controlled studies in which evaluators are themselves being evaluated. General guidelines for increasing reliability in clinical situations include having two observers compare their results, providing careful instructions when a client is asked to monitor his or her own behavior, specifying target behaviors, clearly wording items on self-reports, taking care in the construction of instruments, and thoroughly training observers such as parents or teachers. Reliability of ratings is also likely to be increased by paying closer attention to contextual variables (J. G. Beck, 1994; S. Haynes & O'Brien, 2000).

During the 1960s and 1970s, the validity of various assessment procedures depended primarily on informal content validity. Questionnaires and observational strategies were based on rational considerations regarding what was to be studied and how these measurements were to be made. Few efforts were made to develop empirically derived categories. For example, the assessment of depression might have been based on knowledge about the typical thoughts depressed people seem to have as well as additional variables that seem important regarding social supports and typical antecedent events.

The various areas of observation were selected mostly based on what rationally seemed to be the most critical considerations. Since the early 1980s, increased work has gone into assessing the validity of various methods of behavioral assessment. In general, few validity studies have been performed on behavioral interviews and naturalistic observations, whereas much more has been done on behavioral questionnaires. Most validity studies have been conducted by using relevant outside criteria. Many of the same issues have come up with criterion validity for behavioral assessment as for traditional assessment, including difficulty generalizing to different populations, settings, and methods of administration.

The early behavioral self-report questionnaires relied on content and face validity. Because these questionnaires represented new techniques with a different underlying philosophy, it was believed that they did not have to be judged using the same criteria as the older and more traditional psychometric tests. They were considered to be direct reports of client behaviors and thus little psychometric validity was reported. R. Kaplan and Saccuzzo (1993) criticize this by stating that behavioral self-reports may be "repeating history and reinventing the wheel" (p. 493). They further point out that the "early paper-and-pencil structured personality tests which were finally abandoned in the 1930s are indeed difficult to distinguish from many present-day (behavioral) self-report procedures" (p. 494). The problems of response bias, questionable reliability and validity, no norms, and assumed client truthfulness need to be addressed for any standardized instrument, including behavioral procedures. Many behavioral self-report questionnaires might be best referred to as "idiosyncratic clinical tools" rather than psychometrically sound tests. The familiar argument used for traditional tests is that different assessment procedures serve to provide checks and balances for one another. Although it is often argued that self-reports are supported by other sources of data (direct observation, psychophysiological measurement, internal dialogue), few actual studies on the incremental validity of these procedures have been conducted.

Many behavioral self-report inventories have been developed but have had widely varying degrees of success demonstrating acceptable psychometric qualities. For example, the Rathus Assertiveness Schedule (RAS; Rathus, 1973) has been subjected to traditional psychometric procedures and illustrates the difficulties encountered in this as well as other similar behavioral inventories. Whereas Heimberg, Harrison, Goldberg, Desmarais, and Blue (1979) did not find a very high correspondence between scores on the RAS and observational reports of role plays in an inmate population, the RAS did relate to nonassertiveness in a group of dental students (Rathus, 1972). However, a difficulty with relating assertiveness in role-play situations, which most of the preceding studies used, is that assertiveness in role plays may not relate to assertiveness in naturalistic situations (Bellack, Hersen, & Turner, 1979). Perhaps when subjects are asked to role-play, they can alter their daily level of assertiveness to "act the part" correctly (Higgins, Alonso, & Pendleton, 1979). The RAS similarly has poor criterion validity based on instructor evaluations of observed assertive behavior and grades in a communication course (Tucker, Weaver, Duran, & Redden, 1983). Thus, even though the RAS is a frequently used device in both research and clinical settings, the meaning of the scores might be difficult to evaluate. Other behavioral self-report questionnaires have experienced similar problems.

ASSETS AND LIMITATIONS

Probably the greatest advantage of behavioral assessment is that its practitioners have continually paid attention to its relevance toward treatment. Any measurement of problem behaviors is usually directly tied to how these behaviors can be changed. Furthermore, relevant behaviors are given an empirical functional analysis, which enables clinicians to make baseline measurements of behavior and to assess the antecedents and consequences of these behaviors. An initial functional analysis can then allow clinicians to evaluate whether change has actually occurred during or after treatment. Although many techniques have not been through rigorous traditional validity studies, the emphasis on treatment validity has proven to be attractive to many practitioners. Thus, behavioral assessment is particularly useful for persons using a hypothesis-testing approach and for those who wish to have clear accountability that change has actually taken place. In some situations, however, behavioral assessment can be tied too closely to treatment, particularly in legal assessments or other situations in which assessment and therapy are separate.

A further asset is that behavioral assessment offers a wide range of possible techniques for use in extremely varied contexts. These strategies include self-reports, naturalistic observation, physiological monitoring, structured observation, and self-monitoring. Variations in techniques are consistent with the view that a complete understanding of the person requires multiple modes of assessment. The different assessment modes might involve relevant aspects of person-situation interaction, physiological changes, cognitions, interpersonal relationships, overt behaviors, feelings, imagery, and aspects of the person's larger social system. Many behavioral assessment models organize their approach around stimulus, organism, response, and contingencies (Goldfried, 1982b). Other approaches rely on Lazarus BASIC-ID (Lazarus, 1989) or on Kanfer and Saslow's (1969) functional analysis of behavioral excesses and deficits. These approaches place the person in a much wider context than traditional assessment procedures.

Behavioral assessment is particularly appropriate when a presenting problem is determined primarily by environmental factors. In most cases, a clear, functional relationship (environmental interaction) can be established for disorders such as phobias, marital difficulties, acting out, temper tantrums, and inappropriate classroom behavior. If these behaviors are frequent in occurrence (i.e., smoking, classroom acting out), it is fairly easy to develop a baseline and monitor change. However, quite unique behavior that occurs infrequently (i.e., violation of drugs or firearms in schools) may be much more difficult to measure and monitor (J. Nelson, Roberts, Rutherford, Mathur, & Aaroe, 1999). In addition, behavioral assessment is somewhat less relevant when environmental factors account for a smaller portion of the variance. This may be the case when organic factors may be more important than environmental ones, such as in chronic schizophrenia, certain types of headaches, and head injuries. Although behavioral assessment and intervention can still be effective for such problems, greater difficulties are involved because the environment is relatively less important.

A previously described but extremely important drawback of many behavioral assessment strategies is that they have poor or, at least, untested psychometric properties. Often the attempts to establish reliability and validity have been disappointing. In addition, the accuracy of behavioral observation and interviewing can be distorted

because of observer bias, halo effects, primacy effects, low interobserver agreement, confirmatory bias, and so forth.

Although cognitive behavioral assessment has been given increased importance, in many ways it is contrary to the original spirit of behavioral assessment's emphasis on direct observation. Cognitive assessment is necessarily unobservable and relies on client self-reports. Difficulties might include differences in meaning between the client and the clinician, response biases, assumed honesty of reporting, and assumptions about the equivalence of internal dialogues and their verbal descriptions.

A final limitation of behavioral assessment is that it often requires extensive resources in terms of time, personnel, and equipment. This is particularly true for psychophysiological and observational methods. Surveys of behaviorally oriented professionals have indicated that only a minority of behaviorally oriented clinicians (15% to 25%) use observations in structured or natural settings (Guevremont & Spiegler, 1990). An earlier review found that only 12% of behavioral clinicians endorsed behavioral analysis and only 20% endorsed the use of direct observations (C. Piotrowski & Keller, 1984). As a result, behavioral assessment is frequently limited to interviews and questionnaires (Guevremont & Spiegler, 1990; Sarwer & Sayers, 1998b). An additional drawback is that many behavioral instruments have not been designed to deal with problems frequently encountered in clinical practice, such as dissociative disorders, paranoia, and hypochondriasis.

STRATEGIES OF BEHAVIORAL ASSESSMENT

Behavioral assessment has given rise to numerous and highly varied techniques, many of which are outlined in Hersen and Bellack's (1998) *Dictionary of Behavioral Assessment Techniques* (2nd ed.) and Bellack and Hersen's (1998) *Behavioral Assessment: A Practical Handbook* (4th ed.). For example, Barrios and Hartman (1988) found more than 100 instruments for assessing children's fears and anxieties. Despite this diversity, behavioral assessment strategies can be organized into the general categories of behavioral interviewing, behavioral observation, cognitive behavioral assessment, psychophysiological assessment, and self-report inventories. Each of these approaches varies in the degree to which it emphasizes direct versus indirect measures of the person, as well as in the extent to which it relies on inference. For example, cognitive assessment is more indirect than behavioral observation and relies much more on inferences regarding the degree to which cognitions affect and interact with overt behavior. However, all of these techniques stress developing a functional analysis of behavior through understanding person-environment interaction. They also emphasize that each aspect of assessment is directly relevant to treatment planning and evaluation.

Behavioral Interviewing

Behaviorally oriented interviews generally focus on describing and understanding the relationships between antecedents, behaviors, and consequences (ABC). In addition, a baseline or pretreatment measure of behavior is developed through a systematic consideration of the frequency, intensity, and duration of relevant behaviors. Behaviors might

also be provided with a description of specific behavioral excesses and deficits (Kanfer & Saslow, 1969). Any goal must be capable of being measured and tested in an objective and reliable way, and the client should agree on its relevance (Gresham, 1984). Although the behavioral approach might seem long and involved, the process is simplified by considering only areas that are relevant for treatment.

Despite this emphasis on treatment utility, it is essential to place each aspect of the information derived from a behavioral interview into a wide context. A basic description of a target behavior is simplistic because it does not take into account an interactionist model. For example, a phobia is likely to create difficulties in the client's relationships, which could undermine the person's sense of competence. The person might then react by becoming highly dependent on a primary relationship, reinforcing the sense of helplessness. The helplessness might then reinforce a fear of not being able to cope, which can then interact with and quite possibly exacerbate the phobia. Thus, a complete interview would evaluate not only the existence of and nature of the phobia, but also the effect of the phobia on relationships, work effectiveness, and self-statements. Whereas the earlier behavioral interviews of the 1960s and 1970s often had a narrow focus, current models of behavioral assessment emphasize taking this wider context into consideration.

The general purpose of the behavioral interview is multifaceted. It might help identify relevant target behaviors or select additional behavioral assessment procedures. It also provides an opportunity to obtain informed consent, obtain a history of the problem, identify causal factors related to the presenting problem, develop a functional analysis of the problem behavior, increase client motivation, design intervention programs, and evaluate the effectiveness of previously attempted interventions.

The initial phase of a behavioral interview needs to include many of the elements relevant for traditional interviews. A sufficient degree of rapport needs to be established, a statement needs to be developed of the general and specific purposes of the interview, and a review should be made of the client's relevant history. However, history tends to be de-emphasized in favor of current behaviors because the main cause of client behavior is considered situational rather than historical. Common clinician approaches involve reflective comments, probing, understanding, and expressed empathy. Open-ended questions can be followed up with more direct questioning. However, the extensive use of nondirective techniques is inappropriate in that the clinician must set a clear direction and have the client answer direct questions relevant to a behaviorally oriented approach.

Sometimes clients provide excellent descriptions of their problems and can specify relevant antecedent and consequent conditions. Other clients experience difficulty describing the events surrounding the decision to seek treatment, elaborating on their feelings, stating who referred them, or providing information about how other people might be perceiving their problem. Because a careful behavioral analysis requires a complete description of problem behaviors, the client and therapist must work to establish the extent of the difficulty, where it occurs, when it occurs, and its effects on relationships. Sometimes it is helpful to have the client keep a diary of relevant events and observations. Often clients describe and define their difficulties by relying extensively on general trait descriptions rather than on more behaviorally oriented ones. A behavioral interviewer, then, needs to work with the client to develop specific and easily observable

descriptions. For example, if a client says he or she is a "depressed type of person," this might translate into specific types of behaviors (slow movement, spending too much time in bed, avoiding people, being nonassertive), cognitions (that he or she is no good, a failure), and feelings (hopelessness, apathy). The belief in an underlying permanent trait (illness) needs to be reframed as a group of specific behaviors that are potentially changeable. This reframing process, in itself, is likely to be beneficial to clients because they will be better able to see specific things they can do to change how they feel. Speaking in concrete behavioral terms rather than abstractions is also likely to increase mutual understanding between client and therapist.

A wide-based behavioral assessment should describe not only the specific presenting problem, but also the manner in which the problem has generalized into other areas. In particular, this assessment might involve information about the larger social system. Often, the client's school, work, or family situation can be incorporated into the assessment and treatment program to ensure both immediate and long-term success. In contrast, if a narrow approach to change is taken, the client may attempt to express his or her newly acquired behavior in contexts that are not supportive of it. As a result, previous problem behavior might once again develop to the exclusion of newer, more adaptive behavior. This might be true if the client developed new, effective behaviors that were learned only in the narrow context of the practitioner's office.

An interview should end by providing the client with a summary of the information obtained, an explanation of additional information that is required, and an estimate of the likely success of treatment (Sarwer & Sayers, 1998b). If further information is required, the clinician and client should agree on what is needed and how to obtain it. This might involve instructions for keeping an effective diary, requests for observations from other people, or techniques for self-monitoring of different behaviors. If the interview is a prelude to therapy, additional information should be given about possible strategies for intervention, the length of treatment, possible financial and emotional costs, and assurances that the client will have input into all decisions.

Because most interviews tend to be somewhat informal and haphazard, they frequently provide information with low reliability and validity. For example, T. Wilson and Evans (1983) found a low level of reliability among clinicians trying to specify appropriate target behaviors. Some authors urge that behavioral interviews be structured and standardized. Kratochwill (1985) has suggested that interviews be planned around a four-stage problem-solving process. The first stage is *problem identification* in which the problem is specified and explored, and procedures are established to measure current performance and desired target behaviors. The vague and generalized descriptions that clients typically come in with are developed into specific behavioral descriptions. Next, a *problem analysis* is performed by assessing the client's resources, and by noting the relevant environmental conditions influencing behavior and the context in which the behavior excesses or deficits occur. An interview also needs to establish how a *plan might be implemented,* which would include ongoing procedures for collecting data relevant to the progress of the treatment. Finally, strategies for *treatment evaluation* should be specified by considering the pre- and posttreatment measures to determine whether the intervention was successful.

Witt and Elliott (1983) provide the following somewhat similar outline of expected accomplishments for any behavioral interview:

1. Initially, provide the client with an overview of what needs to be accomplished and why a clear and detailed specification of the problem behavior is important.
2. Identify the target behavior(s) and articulate them in precise behavioral terms.
3. Identify the problem frequency, duration, and intensity ("How many times has it occurred today," "How long has it been going on," etc.).
4. Identify conditions in which the problem occurs in terms of its antecedents, behaviors, and consequences.
5. Identify the desired level of performance and consider an estimate of how realistic this is and possible deadlines.
6. Identify the client's strengths.
7. Identify the procedures for measuring relevant behaviors. What will be recorded, who will record it, how will it be recorded, when and where will it be recorded?
8. Identify how the effectiveness of the program will be evaluated.
9. After completing discussion of the preceding areas, summarize it to ensure that the client understands and agrees.

This outline should not be followed rigidly, but should be used as a general guideline. However, each behavioral assessment should have accomplished all nine areas before completion.

Behavioral Observation

In some cases, the behavioral interview is itself sufficient to obtain an adequate assessment. However, some form of actual behavioral observation is often required before, during, and/or after treatment. The particular method for observing behavior is usually decided on during the initial interview. Whereas the interview is directed primarily toward obtaining verbal information from the client, behavioral observation is used to decide on and actually carry out specific strategies and techniques of measuring the relevant areas of behavior discussed during the interview (see Tryon, 1998). In some cases such as assessing the developmentally disabled, resistant clients, or very young children, behavioral observation may become one of the most important means of assessment. These observations might be made by the professional who is actually conducting the treatment or by someone else who is more involved in the client's life such as a teacher, parent, spouse, or self-monitoring by the client. The most frequent approaches are narrative recording, interval recording, event recording, and ratings recording.

The first behavioral observation task is to select relevant target behaviors, which can vary from a single response set to a larger interactive unit. The target behavior should either involve the problem behavior itself or relate to it in a meaningful way. Decisions must be made regarding the number of behaviors to record and the relative complexity of the recording method. Both the recording method and the target behavior need to be manageable and should avoid being overly complex. The target behavior can best be clarified by beginning with a narrative description of the client's difficulty and then further specified by considering the antecedents and consequences related to the problem behavior.

All behaviors to be measured must have objective, complete definitions that allow clear observations of the measures of the behavior. In particular, the definition should avoid abstract and highly inferential terms, such as apathy or sadness, and instead translate such terms into specific behaviors. Any description of the target behavior should involve an easy-to-read dictionary-type definition, an elaboration of the behavior, and specifications regarding precisely when the behavior occurs, as well as descriptions of borderline examples and clear nonexamples. In measuring behavioral frequencies, the practitioner must clearly define when the behavior begins and ends. This might be easy for measuring the number of cigarettes a person smokes or the number of times a child bangs his or her head, but is more difficult when measuring less clearly defined behaviors, such as the number of aggressive acts a person makes or frequency of nonassertive behaviors. Recordings should also measure the duration of behaviors and their intensity. For example, how hard a child bangs his or her head and the total time engaged in the activity have implications for the urgency and strength of the treatment approach.

The different devices used to make recordings might include various combinations of golf counters, stopwatches, pencil-and-paper forms, or electromechanical devices such as an event recorder with buttons that can be pressed when various categories of behaviors occur. Hand-held computers are becoming more common as well as video and audio recordings.

The settings of behavioral observation can range from those that are natural to those that are highly structured. Natural, or in vivo, settings might include the home, classroom, business, or playground. Observations made from these types of settings are likely to be directly relevant to and reflective of the client's life. Natural settings are most effective when assessing high-frequency behaviors and/or more global behaviors, such as attentional deficits, social withdrawal, or depressive behaviors. They are also useful when measuring the amount of change the client has made following intervention. However, natural settings present difficulties because of the extensive time required to make observations. Furthermore, natural settings are problematic when trying to measure infrequently occurring behaviors (aggression, nonassertiveness) or behaviors that typically occur in the absence of others (fire-setting, suicide). To counter the difficulties inherent in naturalistic observation, practitioners may wish to create structured environments (role plays, work simulations) that elicit specific types of behaviors. Such environments are especially important for infrequent behaviors. However, inferences need to be derived cautiously from observations in these structured or analogue situations, as they may not generalize into the client's actual life.

When clinicians are concerned that observations made by a person outside the client's environment might contaminate the results, they may wish to train persons who are already a part of the client's natural setting, such as parents, teachers, or spouses. This might help prevent subjects from changing their behaviors simply because they are aware that they are being observed (reactivity). These more natural observers can be much less obtrusive than an outside professional. The training of observers needs to include a clear rationale for measuring the behavior with emphasis on making accurate and objective recordings. Observers should memorize the recording code, practice making the recordings, and receive feedback about the relative accuracy of their recordings. Precautions should be taken to avoid observer error, such as

through observer bias, leniency, lapses in concentration, and discussion of data with other observers. Sometimes reliability might be checked by comparing the degree of agreement between different observers rating the same behaviors. Caution should be made when using trained observers because widely varying levels of interobserver agreement have been noted (G. Margolin, Hattem, John, & Yost, 1985).

A system of coding behaviors usually needs to be developed so that recordings are abbreviated and simplified. If too many codes are used, it is difficult for recorders to recall them, especially if behaviors occur in rapid succession. Both the type of recording method (narrative recording, event recording, etc.) and the coding system depend largely on the goals of assessment. A coding system that is clear, simple, and closely connected to the presenting problem is likely to be both useful and reliable. Important considerations in selecting a recording and coding system are the number of times the behavior needs to be observed, the length of observation periods, when to make the recording, the type of recording to be made, and the target behaviors to be recorded. The following sections describe the most frequently used recording systems, along with examples of different methods of coding.

Narrative Recording

Narrative recording requires that the observer simply make note of behaviors of interest. There is little quantification, and the observations can vary in the degree of inferences made. For example, an observer may stick closely to direct descriptions of behavior, such as noting that someone frequently laughs and smiles at his or her friends, or may infer from these behaviors that the client has good peer relations. The primary value of narrative recordings is that they may help define future, more specific areas, which can then be measured in a quantitative manner. Thus, narrative recording is usually a precursor to alternative forms of measurement. It has the advantages of potentially discovering relevant behaviors; it can elaborate on these behaviors; it requires little, if any, equipment; and numerous hypotheses can be generated from the narrative descriptions. Limitations are that it doesn't enable the observer to quantify the observations, it may have questionable validity, and the usefulness of the observations depends largely on the individual skill of the observer.

Interval Recording

A clinician may choose to record whether selected aspects of behavior occur within predetermined intervals. As a result, this technique is also referred to as *time sampling, interval sampling,* or *interval time sampling.* Usually, the intervals vary from 5 to 30 seconds and may be based either on set schedules for each observation period (e.g., every five minutes) or may be selected randomly. Interval recording is most appropriately used when measurements of overt behaviors with moderate frequencies (e.g., once every 5 to 20 seconds) are required and when these behaviors do not have any clear beginning or end. This might include behaviors such as walking, listening, playing, reading, or looking up and down.

When developing a strategy for interval recording, clinicians must decide on the length of time between each observation, the method of recording, and the length of the observation period. This depends largely on the type of behavior. For example, different types of verbal interaction may vary in length and, as such, the observation periods

must be adjusted. Some strategies might require the observer to alternate between recording (e.g., for 10 seconds), then observing (e.g., for 20 seconds), and then going back to recording the observation just made. Cues regarding the beginning and end of each behavior must be specified. The target behaviors for observation are derived from information based on such sources as the initial interview, self-report inventories, narrative observations, and especially from descriptions of the presenting problem. The focus of observation may also vary between different people such as the husband, wife, teacher, child, or client. Sometimes clinicians or researchers arrange to have an outside person observe the same client behaviors. The interrater reliability of the observations can then be established by calculating the percentage of agreement between the two raters (see Tryon, 1998). A representative interval recording chart, with instructions on how to develop such a chart, are provided in Table 4.1.

Interval recording is time efficient, highly focused on specific behaviors, and has the potential to measure almost any behavior. Interval recording is not designed to assess the quality of the target behaviors, however, and can be artificial or may overlook other important behaviors.

Event Recording

Whereas interval recording depends on measurements defined by units of time that are imposed on target behaviors, event recording depends on the occurrence of the behavior itself. The observer must wait for the target behavior to occur, and then record relevant details of the behavior. Examples of behaviors most appropriate for event recording are aggressive actions, greetings, or use of verbal expressions such as assertion or profanity.

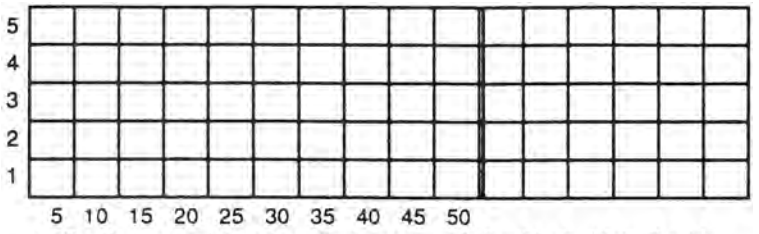
The basic design of event recording systems is to note the behavior's frequency, duration, and intensity, and to record the behavior on such devices as a checklist, golf counter, or hand counter. Although the main emphasis is on quantifying the frequency of responding, its duration also can be measured with a stopwatch. The intensity of the behavior can be noted by simply specifying whether it was slight, moderate, or strong. A representative example of an event-recording chart is included in Table 4.2.

Event recording is especially good for recording behaviors having low frequencies, measuring changes in behaviors over time, and for use in studying many different types of behavior. However, event recording is relatively poor at measuring behaviors that do not have clear beginnings and endings, and presents difficulties in keeping the attention of observers for behaviors of long durations. Because event recording does not provide information regarding sequences of behaviors, it is difficult to make inferences about how and why behaviors occur.

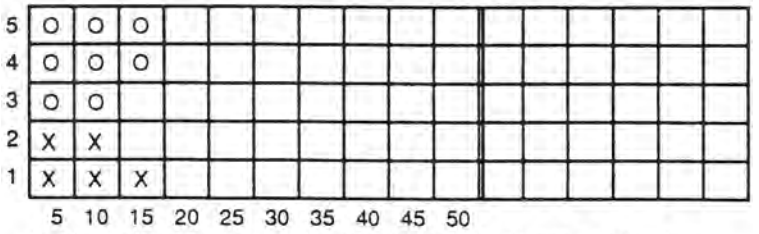
Ratings Recording

Rather than recording direct observations of behaviors, clinicians may wish to obtain general impressions of relevant dimensions of behaviors and have these impressions rated on a checklist or scale. Such measures tend to be more global and may involve more abstract terms, such as the client's level of cooperativeness or ability to maintain self-care. Typically, ratings recordings are made after a period of observation. A typical format might request the evaluator to rate, on a scale from one to five or one to seven, the client's frequency of temper tantrums, quality of peer relations, or conscientiousness. For example, the Motivation Assessment Scale (MAS; Durand, 1990) is a

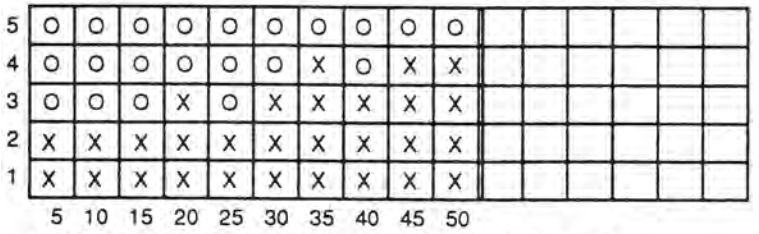
Table 4.1 Example of interval recording



a. Graph paper with series of columns, each five blocks high. Double heavy line marks off 10 columns, for a 50-minute period.



b. Chart after 13 minutes of monitoring pupil's behavior. First two columns are completed and the third is partially completed. If the pupil behaves appropriately during the next (14th) minute, the observer will mark an "X" in the third column just above the other "X." If the pupil misbehaves, the observer will mark an "O" in that column just under the other two "Os."



c. Chart after observer has completed the 50-minute period.

To set up a self-graphing data recording system, start with a piece of graph paper. Mark 2 heavy lines across the paper so that 5 blocks are between the lines. You have now a series of columns, all 5 blocks high. Each block will represent an interval (e.g., minute) of observation time. Mark off the number of 5-block columns needed for the scheduled observation period: a 50-minute period would need 10 columns of 5 blocks; a 30-minute period would need 6 columns; a 45-minute period would need 9 columns; and a 5-minute period would need only 1 column of 5 blocks. For now, let's assume you have scheduled a 50-minute period for your observation, as shown in Table 4.1 a–c. You have marked off 10 columns on your paper, each 5 blocks high, for a total of 50 blocks: 1 block for each minute scheduled.

For each interval (minute) in which the behavior occurs, you will place an "X" in a box. For each interval in which the behavior does not occur, you will place an "O" in a box. Start with the left column and work toward the right. In each column, work from the bottom up with the "Xs," but from the top down with the "O" marks. When the "Xs" and "Os" meet in the middle, the column is filled. Move to the next column to the right and continue: "Xs" from the bottom, "Os" from the top down, until they meet. As you move across the row of 5 columns, the data recorded will automatically form a graph without any extra effort on your part. With this methods, trends in data across the session can be easily identified and shared with school personnel or parents. By focusing on the "Xs" in Table 4.1c, it is clear that the amount of "on task" behavior by the pupil is steadily increasing during the observation session (i.e., there are fewer "Xs" in the first column, and more "Xs" in the later columns).

Source: From "Behavioral Observation for the School Psychologist: Responsive-Discrepancy Model" by G. J. Alessi, 1980, *School Psychology Review,* p. 40. All explanatory material is verbatim from Alessi, 1980.

Table 4.2 Example of event recording within 5-minute intervals

			Intervals in minutes					
Behavior	Totals	Person Observed	5	10	15	20	25	30
Getting out of seat	27	Subject	□	⌐	□	:	.	..
	8	Comparison	..	.	:.	.		.
Request-ing help	5	Subject		.		:.		.
	11	Comparison	..	:.	:.	.	.	.

Table 4.2 illustrates an event recording for two different types of behaviors, the first of which (getting out of seat) the subject's teacher would like to see less of and the second (requesting help) the subject's teacher would like to see more of. In addition to recording the subject's behavior, another student was selected as a basis for comparison. The coding of the number of responses was developed by Tukey, 1977, and uses dots and lines to indicate the number of responses which were made. One dot equals one response. Any number above four responses is indicated by a line connecting two dots. For example, in the first 5-minute block for "getting out of seat," the subject got out of his seat eight times. By noting the increases and decreases in the different recordings, observers can be alerted to possible environmental events that might have caused these changes. In this example, there was both a decrease in "getting out of seat" and an increase in "requesting help" beginning at the 20-minute interval.

16-item questionnaire that evaluates the functional significance of behavior related to the dimensions of sensory, escape/avoidance, social attention, and tangible rewards. Interrater reliability for the MAS ranged between .80 and .95 with test-retest reliability (30 days apart) ranging between .89 and .98. Validity has been supported through means such as determining that teacher's ratings on the MAS predicted students' behavior in analogue situations (Durand, 1990). An example of a completed MAS is illustrated in Table 4.3.

Ratings recordings can potentially be used for a wide variety of behaviors. Other advantages are that the data can be subjected to statistical analysis; the ratings can be made for either individuals or groups; and because of the time efficiency of ratings recordings, they are likely to be cost-effective. Disadvantages include possibly low interrater agreement because of the subjectivity of the ratings; little information regarding antecedent and consequent events; and possibly inaccurate ratings, especially if much time elapses between making the observations and making the ratings.

Cognitive Behavioral Assessment

Over the past 25 years, considerable research has been conducted on understanding the cognitive processes underlying behavior disorders. Relevant areas include the self-statements associated with different disorders, the underlying structure or cognitive organization related to these disorders, differences between cognitive distortions in pathological versus normal behavior, and cognitive alterations that occur during therapy. This research has considerably influenced and altered the nature of behavioral assessment. In particular, researchers have developed specific techniques for assessing cognitive processes, such as having the person think aloud, listing different thoughts, thought sampling at various intervals, and a wide variety of self-statement inventories.

Table 4.3 A completed Motivation Assessment Scale for Bill's object hitting in one-to-one instructional settings

Name Bill Rater Mark Date 7/16/88

Behavior Description Hitting objects (e.g., tops of tables) with his hand.

Setting Description One-to-one instructional settings in school

Instructions: The **Motivation Assessment Scale** is a questionnaire designed to identify those situations in which an individual is likely to behave in certain ways. From this information, more informed decisions can be made concerning the selection of appropriate reinforcers and treatments. To complete the **Motivation Assessment Scale**, select one behavior that is of particular interest. It is important that you identify the behavior *very specifically*. *Aggressive*, for example, is not as good a description as *hits his sister*. Once you have specified the behavior to be rated, read each question carefully and circle the *one* number that best describes your observations of this behavior.

QUESTIONS	ANSWERS						
	Never	Almost Never	Seldom	Half the Time	Usually	Almost Always	Always
	0	1	2	3	4	5	6
1. Would the behavior occur continuously, over and over, if this person was left alone for long periods of time? (For example, several hours.)	(0)						
2. Does the behavior occur following a request to perform a difficult task?						(5)	
3. Does the behavior seem to occur in response to your talking to other persons in the room?		(1)					
4. Does the behavior ever occur to get a toy, food, or activity that this person has been told that he or she can't have?				(3)			
5. Would the behavior occur repeatedly, in the same way, for very long periods of time, if no one was around? (For example, rocking back and forth for over an hour.)	(0)						
6. Does the behavior occur when *any* request is made of this person?					(4)		
7. Does the behavior occur whenever you stop attending to this person?		(1)					
8. Does the behavior occur when you take away a favorite toy, food, or activity?					(4)		
9. Does it appear to you that this person enjoys performing the behavior? (It feels, tastes, looks, smells, and/or sounds pleasing.)		(1)					
10. Does this person seem to do the behavior to upset or annoy you when you are trying to get him or her to do what you ask?						(5)	
11. Does this person seem to do the behavior to upset or annoy you when you are not paying attention to him or her? (For example, if you are sitting in a separate room, interacting with another person.)			(2)				
12. Does the behavior **stop** occurring shortly after you give this person the toy, food or activity he or she has requested?					(4)		
13. When the behavior is occurring, does this person seem calm and unaware of anything else going on around him or her?	(0)						

Table 4.3 *(Continued)*

QUESTIONS	ANSWERS						
14. Does the behavior stop occurring shortly after (one to five minutes) you stop working or making demands of this person?	Never 0	Almost Never 1	Seldom 2	Half the Time 3	Usually 4	Almost Always 5	Always (6)
15. Does this person seem to do the behavior to get you to spend some time with him or her?	Never 0	Almost Never (1)	Seldom 2	Half the Time 3	Usually 4	Almost Always 5	Always 6
16. Does the behavior seem to occur when this person has been told that he or she can't do something he or she had wanted to do?	Never 0	Almost Never 1	Seldom 2	Half the Time (3)	Usually 4	Almost Always 5	Always 6

	Sensory	Escape	Attention	Tangible
	1. 0	2. 5	3. 1	4. 3
	5. 0	6. 4	7. 1	8. 4
	9. 1	10. 5	11. 2	12. 4
	13. 0	14. 6	15. 1	16. 3
Total score =	1	20	5	14
Mean score =	0.25	5.0	1.25	3.5
Relative ranking =	4	1	3	2

Source: From *Severe Behavior Problems: A Functional Communication Training Approach,* pp. 80–82, by V. M. Durand, 1990, New York: Guilford: Copyright © 1990 by Guilford Press. Reprinted by permission.

This internal perspective is quite different from the early emphasis of behavioral assessment, which focused almost exclusively on observable overt behavior. This transition has come about because of persuasive evidence for the relationship between behavior and cognitions (Alloy et al., 1999; Bandura, 1986; Haaga, Dyck, & Ernst, 1991; Schwartz & Garamoni, 1989). Cognitive processes not only change during the course of effective therapy, but may be causally related to both the development as well as the maintenance of different types of disorders (Alloy et al., 1999; Breitholtz et al., 1999; Brewin, 1996; Ingram, Kendall, Siegle, Guarino, & McLaughlin, 1995). Some approaches assume that altering cognitions can be sufficiently powerful to change behaviors. However, there are also a number of significant limitations with cognitive behavioral assessment. All material is necessarily derived from the client's self-reports of his or her internal processes and, as such, may be subject to a number of distortions. Clients can usually recall and describe the results of their cognitive processes, but they have much greater difficulty describing how they arrived at these conclusions. The actual processes may need to be inferred based on complicated analyses of the results derived from intricate assessment strategies. In addition, remembering events seems to be a reconstructive process in which each successive recall can be altered based on the person's needs, biases, and expectations (Henry et al., 1994; Lindsay & Read, 1995; Loftus, 1993). These inherent difficulties have led some traditional behaviorists to question the theoretical and practical appropriateness of cognitive assessment.

A relevant finding is that the popular belief in the "power of positive thinking" is simplistic because it is not a very good predictor of adjustment. What seems more important is the absence of negative statements or, what Kendall and Hollon (1981) have

referred to as "the power of nonnegative thinking." Furthermore, the effect of negative self-talk is greater than the ability of positive thinking to counter negative internal dialogue. As might be expected, gains in therapy have been associated with reductions in negative self-statements (J. G. Beck, 1994). Another issue is that relevant cognitions such as self-efficacy vary across situations. For example, a particular client might have cognitions quite consistent with competency in employment situations yet feel quite incompetent in family or other interpersonal situations. This means that clinicians conducting cognitive and other forms of assessments need to take these contextual variables into consideration (J. G. Beck, 1994).

The two major strategies of cognitive assessment are through various self-report inventories and techniques of recording cognitions. Each of these general strategies has strengths and weaknesses and is appropriate in different situations for different types of clients.

Cognitive Self-Report Inventories

There has been a tremendous expansion in the number and frequency of use of cognitive self-report inventories. Guevremont and Spiegler (1990) noted that they were used nearly as frequently as behavioral interviewing and twice as often as direct observation (Guevremont & Spiegler, 1990). They have the general advantages of having strong face validity and are both easy and inexpensive to administer. However, their psychometric properties vary greatly and many instruments in frequent use are quite poor in this regard. Typically, they involve between 20 and 100 items, with respondents asked to indicate their degree of endorsement of each item on a Likert-type scale. Many of them have been tailored toward specific domains such as depression, fears and anxieties, self-efficacy, imagery, social skills (especially assertiveness), eating disorders, and marital problems. The main domains for cognitive self-report inventories and the most frequently used instruments in these domains are summarized in Table 4.4. It is beyond the scope of this chapter to review them, but useful information can be obtained in Bellack and Hersen (1998) and Hersen and Bellack (1998).

Theories of the cognitive processes of *depression* suggest that it is maintained by characteristic and repetitive thoughts that are self-perpetuating (Alloy et al., 1999). For example, A. T. Beck (1967) listed the cognitions associated with depression as involving *arbitrary inference* (making inferences without substantiating evidence), *selective abstraction* (making a broad judgment based on a minor aspect of an event), *overgeneralization* (extrapolating in an unjustified fashion from a minor event), and *magnification/minimization* (overemphasizing negative events; minimizing positive ones). Although these processes seem to be related to depression, a simple cause-effect model between depression and specific cognitions does not appear to be warranted and further clarification is required (Haaga et al., 1991). The most frequently used inventories to assess depressogenic cognitions are the Dysfunctional Attitudes Scale (A. Weissman & Beck, 1978), the Cognitive Bias Questionnaire (Hammen, 1978; Hammen & Krantz, 1976), Automatic Thoughts Questionnaire (Hollon & Kendall, 1980; Ingram et al., 1995), and Beck Depression Inventory (BDI-II; A. T. Beck et al., 1996). More extensive coverage of the BDI/BDI-II can be found in Chapter 13. In addition, the Attributional Styles Questionnaire (Seligman, Abramson,

Table 4.4 Cognitive self-report measures

Domain	Instruments
Depression	Dysfunctional Attitudes Scale
	Cognitive Bias Questionnaire (child and adult versions)
	Automatic Thoughts Questionnaire
	Beck Depression Inventory
	Attributional Styles Questionnaire
Fears and Anxieties	Social Avoidance and Distress Scale
	Fear of Negative Evaluation Scale
	Social Interaction Self-statement Test
	Irrational Beliefs Test
	Rational Behavior Inventory
	Fear Survey Schedule
Eating Disorders	Eating Attitudes Test
	Bulimia Test-Revised
	Cognitive Error Questionnaire (modified for eating disorders)
Social Skills	Rathus Assertiveness Inventory
	Wolpe-Lazarus Assertion Inventory
	Gambrill Assertion Inventory
	Bakker-Assertiveness Schedule
	Conflict Resolution Inventory
	Survey of Hetereosexual Interactions
	Stanford Shyness Scale
Marital Relationships	Relationship Attribution Measure
	Relationships Beliefs Inventory
	Dyadic Attribution Inventory
	Marital Attitude Survey
	Specific Relationship Standards

Semmel, & von Baeyer, 1979) is sometimes used to better understand the manner in which a client construes the causes for various behaviors, particularly those related to depression (i.e., learned helplessness).

A wide number of measures have been developed related to a person's *fears and anxieties* (see McGlyn & Rose, 1998). The main cognitions that seem to characterize social phobias are interpersonal threat, along with beliefs that positive interpersonal feedback is incorrect (Sewitch & Kirsch, 1984). The importance of a cognitive assessment of social phobias is underscored by research suggesting that cognitive deficits and distortions are more important in causing and maintaining the difficulty than deficits in social skills (Heimberg, 1994). Social phobics are more likely to recall negative information, fear social embarrassment, interpret ambiguous feedback negatively, underestimate their own performance, expect more negative evaluations from others, and have more negative self-statements before interactions (Breitholtz et al., 1999; Cacioppo, Glass, & Merluzzi, 1979; Hope & Heimberg, 1993). Assessment of the relative rate of occurrence of each of these areas can provide specific treatment suggestions regarding

which processes need to be modified. The most frequently used instruments in the cognitive assessment of social phobias are the Social Avoidance and Distress Scale (Watson & Friend, 1969), Fear of Negative Evaluation Scale (FNE; Watson & Friend, 1969), and the Social Interaction Self-Statement Test (Glass, Merluzzi, Biever, & Larsen, 1982). Many of the self-statements described by research on social phobias and measured by tests such as the Social Interaction Self-Statement Test are quite similar to the ones described by A. T. Beck (1967) as characteristic of depression. These similarities raise the still unresolved issue of whether specific irrational beliefs are related to specific disorders, or whether there is a nonspecific (yet generally negative) effect of irrational beliefs (see Heimberg, 1994). Although less work has been done on generalized anxiety, two frequently used tests are the Irrational Beliefs Test (R. Jones, 1969) and the somewhat similar 70-item Rational Behavior Inventory (Shorkey, Reyes, & Whiteman, 1977). Although the many versions of the Fear Survey Schedule (Wolpe & Lang, 1964, 1969, 1977) and the Fear Survey Schedule for Children (Ollendick, 1978, 1983) do not measure specific cognitions related to fear, it is both frequently used and quite useful in detailing the various categories of fear a client might have.

Several strategies have been used in the assessment of *eating disorders* based on the observations that this class of disorders involves considerable cognitive distortions (Mizes & Christiano, 1994). Some authors have taken a previously developed scale such as the Cognitive Error Questionnaire (Lefebvre, 1981) and modified it to evaluate the cognitive distortions specific to eating disorders (Dritschel, Williams, & Cooper, 1991). The Eating Attitudes Test (Garner & Garfinkel, 1979) and the Bulimia Test-Revised (Thelan, Farmer, Wonderlich, & Smith, 1991) both have strong psychometric properties and focus primarily on cognitions related to eating and weight control. A further strategy is to have eating-disordered persons monitor their self-statements in their natural environments (Zotter & Crowther, 1991). The value of such strategies is the indication that cognitive behavioral instruments can be tailored toward specific disorders and the information derived from these strategies has direct relevance for treatment as it provides clinicians with specific cognitions to work with.

The area that has dominated the assessment of *social skills* has been assertiveness. Such assessment typically rates not only cognitions related to assertive behavior, but also specific behaviors and skills. A wide variety of self-report inventories has been developed, including the Wolpe-Lazarus Assertion Inventory (Wolpe & Lazarus, 1966), Gambrill Assertion Inventory (Gambrill & Richey, 1975), Bakker Assertiveness Inventory (Bakker, Bakker-Rabdau, & Breit, 1978), and the Conflict Resolution Inventory (McFall & Lillesand, 1971). However, the RAS (Rathus, 1973) has been the most extensively used, and relevant normative data are available for normal college students (Quillan, Besing, & Dinning, 1977) as well as for psychiatric populations (Rathus & Nevid, 1977). Respondents are requested to rate, on a six-point scale, how descriptive each statement is. A –3 indicates that the statement is "very uncharacteristic of me" and a +3 indicates that it is "very characteristic." In addition to the original 30-item schedule, two other versions have been developed for special populations. The modified RAS (MRAS; Del Greco, Breitbach, & McCarthy, 1981) was developed for young adolescents. A simplified version is available that requires a minimum 6th-grade reading level in contrast to the 10th-grade reading level required for the regular version (SRAS; McCormick, 1984). Additional, nonassertiveness social skills inventories

include the Survey of Heterosexual Interactions (Twentyman & McFall, 1975) and the Stanford Shyness Survey (Zimbardo, 1977).

Assessing *marital relationships* involves gathering information about a wide range of behaviors with a particular focus on the strengths and weaknesses of the relationship, goals for change, and attempts they have made to change in the past. Much of this information can and should be obtained through a careful interview. Areas related to cognitive assessment are the differing perceptions of each spouse, the perceived causes (attributions) for why the persons act in certain ways, expectations for future behavior, assumptions about relationships (roles, scripts), and standards by which the relationship is judged (Sayers & Sarwer, 1998a). Many of these areas can be evaluated through the use of cognitive self-report inventories. Some of the more frequent and well-researched instruments are the Relationship Attribution Measure (Fincham & Bradbury, 1992), Relationships Beliefs Inventory (Eidelstein & Epstein, 1982), Dyadic Attribution Inventory (Baucom, Sayers, & Duhe, 1989), Marital Attitude Survey (Pretzer, Epstein, & Fleming, 1992), and Specific Relationship Standards (Baucom, Epstein, Rankin, & Burnett, 1996).

Self-efficacy has received considerable interest, particularly because it has been related to a variety of different predictions relevant to treatment (Bandura, 1986). Assessment is usually accomplished by simply having clients rate the degree to which they believe they are able to accomplish a certain skill or goal (i.e., stop smoking). Useful distinctions should be made between the level of strength of self-efficacy and its generalizability from one situation to the next. Because some question exists regarding the degree to which self-efficacy can be related from one situation to the next, specific measurements are often used for different areas (depression, assertion, smoking, etc.). A person having a high level of self-efficacy is likely to have positive expectations about his or her effectiveness to judge and deal effectively with situations. Self-efficacy is developed as a result of the attainments someone has achieved in the past, vicarious (observational) experiences, verbal persuasion, and physiological states. An assessment of self-efficacy is especially important in understanding the antecedent and retrospective accounts of the effect and quality of the behavior. The relative level of self-efficacy has been found to predict a wide number of variables, including general therapy outcome (O'Leary, 1985), the prediction of success in the treatment of smoking (J. Baer, Holt, & Lichtenstein, 1986; J. Baer & Lichtenstein, 1988; DiClemente, 1986), and relapse rate from self-regulatory training (J. Carlson, 1982).

An area needing further development is the *clinical assessment of imagery.* It has frequently been observed that a person's presenting problem is significantly related to his or her fantasies or daydreams and different dreaming states. A depressed person may continually repeat images of being criticized, the anxious person might replay scenes of danger, and the paranoid might frequently review images of persecution. Knowing a person's relative ability to produce and control images may be important in predicting response to treatment that requires the formation of images such as systematic desensitization, covert desensitization, covert aversive conditioning, and certain types of relaxation procedures. Extensive experimental work has been conducted on imagery in areas such as the different dimensions of imagery (C. Parks, 1982), differences between waking and nonwaking imagery (Cartwright, 1986), and the effects of conscious and unconscious images on behavior (Horowitz, 1985). However, little material has been published regarding the clinical assessment of imagery. Of the studies

that have been published, most have related to measures of imagery ability rather than to the effect of clinically relevant images on the person. Persons wishing to assess both clinical imagery and other aspects of cognitions might use one or several of the following strategies that have been developed to assess cognitions.

Recording Cognitions

In addition to the many self-report inventories available, a number of strategies have been developed for recording cognitions in a less-structured manner. C. Parks and Hollon (1988) have listed and summarized the following methods used by previous researchers:

Thinking Aloud. Clients are requested to verbalize their ongoing thoughts, with these verbalizations usually extending for 5 to 10 minutes (Lodge et al., 2000). A similar technique is free association, in which the client is asked to simply say whatever comes to mind rather than report on his or her ongoing inner thoughts. A potential problem is that the procedure may feel unnatural and, therefore, provide a sample different from normally occurring internal thoughts. Also, the client may have no opportunity to verbalize competing thoughts with the result that the reported thoughts will most likely be a limited portion of the total cognitions. In addition, clients may not report everything honestly. A factor that is likely to make the verbally reported thoughts different from actual ongoing processes is that, typically, people change the topic of ongoing internal dialogues every 5 to 6 seconds, whereas verbal reports of these dialogues may have topic changes only on the average of every 30 seconds.

Private Speech. Sometimes, children's cognitions can be assessed by paying close attention to barely audible speech they make while engaged in various activities. It is believed that these private verbalizations are closely aligned to inner thoughts.

Articulated Thoughts. Clinicians may wish to create structured situations or simulations that parallel the problems the client reports. For example, a situation may be created that demands the client to be assertive or be exposed to criticism or phobic stimuli. The person can then be asked to articulate the thoughts he or she is experiencing during these situations. Typical thoughts can be noted and inferences made regarding how they relate to the problem behaviors.

Production Methods. Instead of asking clients to articulate their thoughts during a simulation, an actual naturalistic situation (criticism, phobic stimuli, etc.) can occur, with clients then noting and recording the typical thoughts they have related to these situations. As such, these methods might also be referred to as *in vivo self-reports.*

Endorsement Method. The client might be presented with either a standardized (e.g., Irrational Beliefs Test, Cognitive Bias Questionnaire) or an informally developed list of items and then be requested to rate frequency of occurrence, strength of belief, and how the item might be uniquely represented in the person's cognitions. These items might include ratings of the frequency of such thoughts as "What's the use" or "I can't do anything right." Potential difficulties with this technique are the

effects of the demand characteristics of the situation and social desirability. An underlying and questionable assumption behind the technique is that the relevant cognitions are in the client's conscious awareness.

Thought Listing. Instead of developing a continuous description of ongoing thoughts, clients might be asked simply to summarize their relevant thoughts. These thoughts might be elicited by a specific stimulus, problem area, or by merely attending to or anticipating a stimulus.

Thought Sampling. A sample of a person's thoughts might be obtained by setting a prompt (e.g., a beep on a timer), then having the client describe the thoughts he or she was having just before the interruption by the prompt.

Event Recording. The client might be asked to wait until a relevant event occurs (e.g., hand washing for an obsessive-compulsive), at which point, the thoughts related to these events are written down. Instead of merely waiting for a problem or spontaneously occurring behavior, a client might also be asked to describe the thoughts related to the expression of new and desired behaviors, such as assertion. The relevant thoughts about these behaviors might then be used to increase the likelihood of their continued occurrence.

Psychophysiological Assessment

A complete understanding of the person involves an assessment of not only behavioral, affective, and cognitive modes, but also of the ways these interact with and are dependent on physiological functioning. Such psychophysiological assessments have recently become easier to make because of increased interest and knowledge regarding instrumentation (electronics, computers), operant conditioning of behaviors that at one time were considered involuntary, physiological and neurochemical aspects of behavior, and behavioral medicine (S. Haynes, 1991; Sturgis & Gramling, 1998). The most frequently assessed physiological responses are heart rate, blood pressure, skin temperature, muscle tension, vasodilation, galvanic skin response (GSR), and brain activity as measured by electroencephalograms (EEGs). By quantifying data gathered through these areas, psychological problems can be translated into more precise physiological indices.

One of the first relevant studies to relate psychological and physiological modes indicated that fear and anger had different physiological responses in blood pressure and skin conductance (Ax, 1953). This result suggested that these and other psychological variables might be measured in ways other than through self-report inventories. More recently, it has been found that persons scoring high on psychological indices of intelligence had relatively small pupillary dilations, lower heart-rate variability, and less skin conductance when asked to perform tasks (Geiselman, Woodward, & Beatty, 1982). This suggests not only that persons with higher intelligence require less effort to complete a task but that, potentially, intellectual assessment might increasingly be based on psychophysiological measurement. A further representative area of research has involved the relationship between different personality variables and psychophysiological measurement (Iacono, 1991). Persons with schizophrenia (when unmedicated) and persons with anxiety disorders have been found to have a relatively higher level of sympathetic responsiveness compared with parasympathetic responsiveness. In contrast,

antisocial personalities are characterized by parasympathetic dominance and low levels of sympathetic responsiveness (Iacono, 1991). Physiological indicators to detect lying, while still in extensive use, have not been found to have adequate psychometric properties (Saxe, Dougherty, & Cross, 1985). Greater promise has been demonstrated differentiating true from faked memory loss using event-related potentials (J. Allen, Iacono, & Danielson, 1992). While most of the previously mentioned studies represent very general correlations among such variables as emotions, intelligence, and behavioral disorders, they show considerable potential for future assessment should these measures become more refined. Physiological baseline measures for an area such as anxiety can and have been used to monitor the effectiveness of treatment for social phobias, generalized anxiety disorders, and obsessive-compulsive disorders (Turpin, 1991).

In addition to the usual knowledge relating to psychological assessment, clinicians who obtain and interpret psychophysiological data must have knowledge in anatomy, electronics, and the physiology of cardiovascular, musculoskeletal, neurological, respiratory, electrodermal, ocular, and gastrointestinal response systems. This extensive background is particularly important because instrumentation presents a number of special problems. A variety of confounding factors may be present, such as the effect of slowing respiratory rate to alter cardiac output or the effect of eye roll on measured brain activity. Filters might be necessary to exclude noise in the system. The techniques are also intrusive, thereby making the situation artificial. As a result, it may not be correct to generalize to outside aspects of the client's life or between different response modes. A wide variety of difficulties may arise regarding meaningful psychological interpretations based on the physiological data. In the future, the development of better instruments and improved methods of computer analysis are likely to increase the utility of psychophysiological assessment and overcome many of these difficulties.

RECOMMENDED READING

Bellack, A. S., & Hersen, M. (Eds.). (1998). *Behavioral assessment: A practical handbook* (4th ed.). New York: Pergamon Press.

Durand, V. M. (1990). *Severe behavior problems: A functional communication training approach*. New York: Guilford Press.

Hersen, M., & Bellack, A. S. (Eds.). (1998). *Dictionary of behavioral assessment techniques* (2nd ed.). New York: Pergamon Press.

Chapter 5

WECHSLER INTELLIGENCE SCALES

The Wechsler intelligence scales are individually administered, composite intelligence tests in a battery format. They assess different areas of intellectual abilities and create a situation in which aspects of personality can be observed. Each of the different versions of the Wechsler intelligence scales provides three different IQ scores: an overall or Full Scale IQ, a Verbal IQ, and a Performance IQ. More specific factor or index scores also can be calculated using various combinations of subtests. The Wechsler intelligence scales are considered to be among the best of all psychological tests because they have sound psychometric properties and produce information relevant to practitioners. As a result, they have become the most frequently used tests in clinical practice (Camara et al., 2000; Watkins et al., 1995).

TESTING OF INTELLIGENCE: PRO AND CON

The testing of intelligence has had a consistent history of misunderstanding, controversy, and occasional misuse (D. Flanagan et al., 1997; Mackintosh, 1998; Weinberg, 1989). Criticisms have ranged from moral indictments against labeling individuals, to cultural bias, and even to accusations of flagrant abuse of test scores. Although valid criticisms can be made against testing intelligence, such procedures also have a number of advantages.

One of the main assets of intelligence tests is their accuracy in predicting future behavior. Initially, Binet was able to achieve a certain degree of predictive success with his scales, and, since that time, test procedures have become progressively more refined and accurate. More recent studies provide ample support that intelligence tests can predict an extremely wide number of variables. In particular, IQ tests are excellent predictors of academic achievement (see R. Gregory, 1999; Mackintosh, 1998; Neisser et al., 1996), occupational performance (J. Hunter & Schmidt, 1996; F. Schmidt & Hunter, 1998; R. Wagner, 1997), and are sensitive to the presence of neuropsychological deficit (Groth-Marnat, 2002; Groth-Marnat, Gallagher, Hale, & Kaplan, 2000; Lezak, 1995; Reitan & Wolfson, 1993). However, certain liabilities are also associated with these successes. First, intelligence tests can be used to classify children into stereotyped categories, which limit their freedom to choose fields of study. Furthermore, IQ tests are quite limited in predicting nontest or nonacademic activity, yet they are sometimes incorrectly used to make these inferences (Snyderman & Rothman, 1987; Sternberg, 1999). It should also be stressed that intelligence tests are measures of a person's present level of functioning and, as such, are best used for making short-term predictions.

Long-term predictions, although attempted frequently, are less accurate because there are many uncontrolled, influencing variables. Similarly, even short-term academic placements made solely on the basis of an IQ score have a high chance of failure because all the variables that may be crucial for success are not and cannot be measured by an intelligence test. It can sometimes be tempting for test users to extend the meaning of test scores beyond their intended scope, especially in relation to the predictions they can realistically be expected to make.

In addition to predicting academic achievement, IQ scores have also been correlated with occupation, ranging from highly trained professionals with mean IQs of 125, to unskilled workers with mean IQs of 87 (Reynolds, Chastain, Kaufman, & McLean, 1987). Correlations between job proficiency and general intelligence have been highest in predicting relatively more complex jobs rather than less demanding occupations. J. Hunter (1986) reported moderately high correlations between general intelligence and success for managers (.53), salespersons (.61), and clerks (.54). For intellectually demanding tasks, nearly half the variance related to performance criteria can be accounted for by general intelligence (F. Schmidt, Ones, & Hunter, 1992). The use of intelligence tests for personnel selection has demonstrated financial efficacy for organizations (F. Schmidt & Hunter, 1998). In addition, the accuracy of using IQ tests can be incrementally increased by combining the results with integrity tests, work samples, and structured interviews (F. Schmidt & Hunter, 1998).

Another important asset of intelligence tests, particularly the WAIS-III and WISC-III, is that they provide valuable information about a person's cognitive strengths and weaknesses. They are standardized procedures whereby a person's performance in various areas can be compared with that of age-related peers. In addition, useful comparisons can be made regarding a person's pattern of strengths and weaknesses. The WAIS-III, WISC-III, and other individually administered tests provide the examiner with a structured interview in which a variety of tasks can be used to observe the unique and personal ways the examinee approaches cognitive tasks. Through a client's interactions with both the examiner and the test materials, an initial impression can be made of the individual's self-esteem, behavioral idiosyncrasies, anxiety, social skills, and motivation, while also obtaining a specific picture of intellectual functioning.

Intelligence tests often provide clinicians, educators, and researchers with baseline measures for use in determining either the degree of change that has occurred in an individual over time or how an individual compares with other persons in a particular area or ability. This may have important implications for evaluating the effectiveness of an educational program or for assessing the changing abilities of a specific student. In cases involving recovery from a head injury or readjustment following neurosurgery, it may be extremely helpful for clinicians to measure and follow the cognitive changes that occur in a patient. Furthermore, IQ assessments may be important in researching and understanding more adequately the effect on cognitive functioning of environmental variables, such as educational programs, family background, and nutrition. Thus, these assessments can provide useful information about cultural, biological, maturational, or treatment-related differences among individuals.

A criticism leveled at intelligence tests is that almost all have an inherent bias toward emphasizing convergent, analytical, and scientific modes of thought. Thus, a person who emphasizes divergent, artistic, and imaginative modes of thought may be at a

distinct disadvantage. Some critics have even stressed that the current approach to intelligence testing has become a social mechanism used by people with similar values to pass on educational advantages to children who resemble themselves. Not only might IQ tests tend to place creative individuals at a disadvantage but also they are limited in assessing nonacademically oriented intellectual abilities (Gardner, 1999; Snyderman & Rothman, 1987). Thus, social acumen, success in dealing with people, the ability to handle the concrete realities of the individual's daily world, social fluency, and specific tasks, such as purchasing merchandise, are not measured by any intelligence test (Greenspan & Driscoll, 1997; Sternberg, 1999). More succinctly, people are capable of many more cognitive abilities than can possibly be measured on an intelligence test.

Misunderstanding and potential misuse of intelligence tests frequently occur when scores are treated as measures of innate capacity. The IQ is not a measure of an innate fixed ability, nor is it representative of all problem-solving situations. It is a specific and limited sample, made at a certain point in time, of abilities that are susceptible to change because of a variety of circumstances. It reflects, to a large extent, the richness of an individual's past experiences. Although interpretation guidelines are quite clear in pointing out the limited nature of a test score, there is a tendency to look at test results as absolute facts reflecting permanent characteristics in an individual. People often want a quick, easy, and reductionist method to quantify, understand, and assess cognitive abilities, and the IQ score has become the most widely misused test score to fill this need.

An important limitation of intelligence tests is that, for the most part, they are not concerned with the underlying processes involved in problem solving. They focus on the final product or outcome rather than on the steps involved in reaching the outcome. They look at the "what" rather than the "how" (Embretson, 1986; E. Kaplan et al., 1999; Milberg et al., 1996). Thus, a low score on Arithmetic might result from poor attention, difficulty understanding the examiner because of disturbances in comprehension, or low educational attainment. The extreme example of this "end product" emphasis is the global IQ score. When the examiner looks at the myriad assortment of intellectual abilities as a global ability, the complexity of cognitive functioning may be simplified to the point of being almost useless. The practitioner can apply labels quickly and easily, without attempting to examine the specific strengths and weaknesses that might make precise therapeutic interventions or knowledgeable recommendations possible. Such thinking detracts significantly from the search for a wider, more precise, and more process-oriented understanding of mental abilities.

A further concern about intelligence tests involves their limited usefulness in assessing minority groups with divergent cultural backgrounds. It has been stated that intelligence-test content is biased in favor of European American, middle-class values. Critics stress that minorities tend to be at a disadvantage when taking the tests because of deficiencies in motivation, lack of practice, lack of familiarity with culturally loaded items, and difficulties in establishing rapport. Numerous arguments against using intelligence tests for the assessment and placement of minorities have culminated in legal restrictions on the use of IQ scores. However, traditional defenses of IQ scores suggest that they are less biased than has been accused. For example, the removal of biased items has done little to alter overall test scores, and IQs still provide mostly accurate predictions for many minorities (see Chapter 2 for a further discussion). The issue

has certainly not been resolved, but clinicians should continue to be aware of this dilemma, pay attention to subgroup norms, and interpret minority group IQ scores cautiously (see Lopez, 1997). Finally, many people feel that their IQs are deeply personal pieces of information. They would prefer that others, even a psychologist who is expected to observe confidentiality, not be allowed access to this information. This problem is further compounded when IQ scores might be given to several different persons, such as during legal proceedings or personnel selection.

Intelligence tests provide a number of useful and well-respected functions. They can adequately predict short-term scholastic performance; assess an individual's relative strengths and weaknesses; predict occupational achievement; reveal important personality variables; and permit the researcher, educator, or clinician to trace possible changes in an individual or population. However, these assets are helpful only if the limitations of intelligence tests are adequately understood and appropriately taken into consideration. They are limited in predicting certain aspects of occupational success and nonacademic skills, such as creativity, motivational level, social acumen, and success in dealing with people. Furthermore, IQ scores are not measures of an innate, fixed ability, and their use in classifying minority groups has been questioned. Finally, there has been an overemphasis on understanding the end product of cognitive functioning and a relative neglect in appreciating underlying cognitive processes.

HISTORY AND DEVELOPMENT

During the 1930s, Wechsler began studying a number of standardized tests and selected 11 different subtests to form his initial battery. His search for subtests was in part guided by his conception that intelligence is global in nature and represents a part of the greater whole of personality. Several of his subtests were derived from portions of the 1937 revision of the Stanford-Binet (Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary). The remaining subtests came from the Army Group Examinations (Picture Arrangement), Koh's Block Design (Block Design), Army Alpha (Information, Comprehension), Army Beta (Digit Symbol-Coding), Healy Picture Completion (Picture Completion), and the Pinther-Paterson Test (Object Assembly). These subtests were combined and published in 1939 as the Wechsler-Bellevue Intelligence Scale. The Wechsler-Bellevue had a number of technical deficiencies primarily related to both the reliability of the subtests and the size and representativeness of the normative sample. Thus, it was revised to form the Wechsler Adult Intelligence Scale (WAIS) in 1955, and another revised edition (WAIS-R) was published in 1981. The 1981 revision was based on 1,880 individuals who were generally representative of the 1970 census and categorized into nine different age groups.

The Wechsler Adult Intelligence Scale-III (WAIS-III) became available in August 1997 and was developed to revise the earlier (1981) WAIS-R. The primary reason for the revision was to update the norms. Additional reasons included extending the age range, modifying items, developing a higher IQ "ceiling" and "floor," decreased reliance on timed performance, developing index/factor scores, creating linkages to other measures of cognitive functioning/achievement, and extensive testing of reliability validity. Despite these changes, many of the traditional features of the WAIS-R

were maintained, including the six Verbal subtests and the five Performance subtests. This still enables practitioners to calculate the Full Scale, Verbal, and Performance IQs. An added feature of the WAIS-III is the inclusion of three new subtests, which enables the calculation of four index scores. Thus, the WAIS-III is not merely a renormed "facelift," but also enables the clinician to do more with the different test scores. This might involve being able to assess persons with either greater age or IQ ranges as well as linking scores with the Wechsler Memory Scales or calculating both IQ and index/factor scores.

The above additions and arrangement of subtests represent the most obvious changes on the WAIS-III. Although not as obvious, its restandardization also represents a major development. The sample was composed of 2,450 adults between the ages of 16 and 89. Each of the 13 age groups was composed of 200 participants with the exception of the 80 to 84 and 85 to 89 age groups, which contained 150 and 100 participants, respectively. Gender and ethnicity closely corresponded to the 1995 U.S. Census data. This included a slightly greater number of women than men at the higher age levels to represent the greater proportion of females in this group. European Americans, African Americans, and Hispanics were also represented in each age band according to the 1995 Census data. The sample was selected from all geographical regions in the United States and stratified to represent the different educational levels in each age group.

The original Wechsler-Bellevue Scale was developed for adults, but in 1949, Wechsler developed the Wechsler Intelligence Scale for Children (WISC) so that children from the age of 5 years 0 months could be assessed in a similar manner. Easier items, designed for children, were added to the original scales and standardized on 2,200 European American boys and girls selected to be representative of the 1940 census. However, some evidence shows that Wechsler's sample may have been overrepresentative of children in the middle and upper socioeconomic levels. Thus, ethnic minorities and children from lower socioeconomic levels may have been penalized when compared with the normative group. The WISC was revised in 1974 and standardized on a new sample that was more accurately representative of children in the United States. The WISC-III (Wechsler, 1991) was released in 1991 with the major change being the inclusion of four factor/index scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed). The new Processing Speed factor has involved the inclusion of the new subtest of Symbol Search along with the older Coding subtest. As with the earlier WISC-R, the standardization and reliability are excellent. The scales were standardized on 2,200 children between the ages of 6 and 16 who closely matched the 1988 census. The sample consisted of 100 boys and 100 girls for each of the different age groups. The new materials are colorful, contemporary, and easy to administer (see review by Little, 1992). The WISC-IV is anticipated to become available in 2003/2004.

In 1967, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) was first published for the assessment of children between the ages of 4 and 6 years 6 months. Just as the WISC is a downward extension of the WAIS, so the WPPSI is generally a downward extension of the WISC in which easier but similar items are used. Although most of the scales are similar in form and content to the WISC, a number of them are unique to the WPPSI. The WPPSI was revised in 1989 (WPPSI-R; Wechsler, 1989) and again in 2002 (WPPSI-III; Psychological Corporation, 2002).

RELIABILITY AND VALIDITY

WAIS-III Reliability and Validity

The reliabilities for the WAIS-III are generally quite high (Psychological Corporation, 1997). Areas of note are that average split-half reliability for the Full Scale IQ (FSIQ) is .98, Verbal IQ (VIQ) is .97, and Performance IQ (PIQ) is .94. The Index reliabilities were similarly quite high with a Verbal Comprehension of .96, Perceptual Organization of .94, Working Memory of .93, and a Processing Speed reliability of .87. The somewhat lower reliability for the Processing Speed Index is primarily because only two subtests (Digit Symbol-Coding and Symbol Search) were used to calculate this index. It should also be noted that, since these two subtests are speeded tests, it was not appropriate to use split-half reliability, and test-retest reliability was calculated instead. Reliabilities for the individual subtests were, as expected, somewhat lower. The highest reliabilities were for Vocabulary (.93) and Information (.91) with the lowest for Object Assembly (.70) and Picture Arrangement (.74). Average subtest test-retest reliabilities over a 2- to 12-week interval ($M = 34.6$ days) were generally comparable, although slightly lower, than the above split-half reliabilities.

While the above test-retest reliabilities indicate a high degree of temporal stability, there is still some degree of improvement on retesting because of practice effects. The Full Scale IQ was found to increase by 4.5 points, the Verbal IQ increased 2.4 points, and the Performance Scale increased a much larger 6.5 points. These increases are not only statistically significant but may have clinical significance when making inferences about the extent to which real improvement/deterioration has occurred for a particular client. This can be crucial when interpreting either specific performance subtests or scores derived from the performance subtests (i.e., Performance IQ, Processing Speed). Thus, a client who has a Performance IQ increase of 6 points on retesting may not really be improving in his or her everyday functions but will merely be demonstrating practice effects. Indeed, a difference of 15 points (for ages 16 to 54) would be required to infer that there has been an actual improvement in abilities (Kaufman & Lichtenberger, 2002). Research with the WAIS-R indicates that these practice effects can occur up to nine months later even among head-injured patients (see p. 148). However, retest gains have also been found to diminish with age (J. Ryan, Paolo, & Brungardt, 1990).

Because extensive validity studies exist for the WAIS-R, one of the most important initial steps in WAIS-III validation was to determine the comparability between the two tests. Comparability would be expected given that the two versions share 70% of their items. As expected, correlations were found to be a quite high .94, .86, and .93 for the Verbal, Performance, and Full Scale IQs, respectively (Psychological Corporation, 1997). This suggests that the WAIS-III measures essentially the same constructs as the WAIS-R. Noteworthy high correlations between the different subtests were .90 for Vocabulary, .83 for Information, and .82 for Digit Span. In contrast, relatively low correlations were found for Picture Completion (.50), Picture Arrangement (.63), and Object Assembly (.69). Correlations between the WAIS-III and WISC-III for a group of 16-year-olds were also quite high (VIQ = .88, PIQ = .78, FSIQ = .88). The index scores were somewhat more variable (Verbal Index = .87, Perceptual Organization Index = .74, Working

Memory=.50, Processing Speed=.79). The low correlation for Working Memory is most likely because the WAIS-III includes the new Letter-Number Sequencing subtest. In contrast, the WISC-III uses only Arithmetic and Digit Span to determine the Working Memory Index. These sets of correlations indicate a mostly high level of correspondence between the WAIS-III and WAIS-R as well as the WAIS-III and WISC-III.

The WAIS-III has also been found to correlate highly with several standard ability measures (Psychological Corporation, 1997). The Standard Progressive Matrices is an untimed, nonverbal test and, as such, the WAIS-III correlations between the Performance IQ and Perceptual Organization Index were moderately high (.79 and .65, respectively). In contrast (and consistent with the construct that the Standard Progressive Matrices is both untimed and nonverbal), the correlation with the Processing Speed Index was low (.25). The correlation between the WAIS-III and Stanford-Binet IV was .88. Further, high to moderate correlations (typically in the high .60s to .70s) were found between the WAIS-III and the Wechsler Individual Achievement Test (Psychological Corporation, 1997). While beyond the scope of this review, correlations have also supported expected associations with measures of attention and concentration, memory, language, fine motor speed/dexterity, spatial processing, and executive functioning (Psychological Corporation, 1997).

Because the Wechsler Memory Scales-III (WMS-III) and WAIS-III have been more closely linked, it is important to evaluate the extent and manner in which they were related (Psychological Corporation, 1997). Correlations between the WAIS-III IQ/Index scores and WMS-III Index scores have generally ranged from .33 to .77 (Psychological Corporation, 1997, p. 124). The VIQ was found to correlate moderately with both the WMS-III Verbal Memory Index (.71) and Visual Memory Index (.73). However, somewhat low correlations were found between the WAIS-III PIQ and WMS-III Visual Immediate (.39) and Visual Delayed (.44) scores. The strongest correlation was between WAIS-III Working Memory and WMS-III Working Memory (.82), which is expected because they share the Digit Span and Letter-Numbering subtests (Psychological Corporation, 1997, p. 93). This pattern of correlations between the WAIS-III and standard tests of intelligence, achievement, and memory provides support for the convergent and divergent validity of the WAIS-III.

Factor analysis of the WAIS-III has supported the presence of *g* in that most subtests correlate with each other, as well as with the FSIQ at least to a moderate extent (Caruso & Cliff, 1999; Psychological Corporation, 1997). Dividing subtests into four Indexes is supported by current theories of intelligence as well as factor analytic procedures (Saklofske, Hildebrand, & Gorsuch, 2000; Wechsler, 1997a) although the fourth Processing Speed factor was found to be relatively weak (Ward, Ryan, & Axelrod, 2000). Despite this, the Processing Speed factor/index has been found to be particularly sensitive to brain dysfunction (K. Hawkins, 1998). In contrast to a clear four-factor solution, Caruso and Cliff (1999) stress that the two most reliable factors were related to crystallized intelligence (*Gc;* composed primarily of Vocabulary and Information) and fluid intelligence (*Gf;* composed primarily of Digit Span and Matrix Reasoning).

A variety of clinical populations has patterns of deficits in learning, cognition, and memory. It would thus be expected that the WAIS-III would be sensitive to these patterns. This was somewhat supported in that the mean WAIS-III IQ and index scores for Alzheimer's disease patients were lower than expected when compared with their

age-related peers. Comparisons among the index scores indicated differential cognitive abilities in that the mean Verbal Comprehension Index was relatively higher (93.0) than either the Processing Speed ($M = 79.6$) or Perceptual Organization ($M = 84.8$) Index. However, it would have been expected that the Working Memory Index would have been somewhat lower than the mean of $M = 87.2$ given the considerable memory complaints among this population. A variety of other neurological disorders (Huntington's disease, Parkinson's disease, traumatic brain injury) found somewhat similar patterns to those with Alzheimer's disease in that verbal abilities were relatively spared (relatively higher VIQ and Verbal Comprehension Index) whereas Processing Speed was lowest. This indicates that the WAIS-III is sensitive to the difficulties these patient populations have with rapidly processing and consolidating information.

Whereas the mean IQ scores for clients diagnosed with attention-deficit hyperactivity disorder (ADHD) did not differ from the standardization sample, the mean Working Memory Index scores were 8.3 points lower than their Verbal Comprehension Index scores (Psychological Corporation, 1997). Similarly, subjects diagnosed with learning disabilities were found to have IQ scores within the normal range (Psychological Corporation, 1997). However, pronounced discrepancies on Index scores were found. Mean Verbal Comprehension scores were 7 points higher than Working Memory scores for reading-disabled subjects and 13 points higher for math-disabled subjects. A subgroup (47.7%) of persons with reading disabilities had at least a 15-point higher mean Verbal Comprehension than Working Memory scores. Discrepancies were further reflected in that mean Perceptual Organization scores were 7 points higher than Processing Speed scores for both math and reading-disabled groups. The ACID profile (lower **A**rithmetic, **C**oding, **I**nformation, **D**igit Span) was also found in that 24% of learning disabled subjects expressed a partial ACID profile and 6.5% expressed a pronounced ACID profile. However, the Verbal Comprehension/Working Memory and Perceptual Organization/Processing Speed discrepancies seemed to more strongly reflect the patterns of cognitive strengths and weaknesses than the ACID profile. This data indicates that the WAIS-III accurately reflected the patterns of deficits related to known characteristics of various clinical and psychoeducational groups.

WISC-III Reliability and Validity

The WISC-III has generally excellent reliability. The average WISC-III internal consistency reported by Wechsler (1991) across all 11 age groups was .96 for the Full Scale IQ, .95 for the Verbal Scale, and .91 for the Performance Scale. Internal consistency for the specific subtests was far more variable, ranging from a low for Object Assembly of .69 to a high of .87 for Vocabulary. The average reliabilities for Verbal subtests ranged between .77 and .87 (*Mdn* $r = .83$), while the Performance subtests were somewhat lower, ranging between .69 and .89 (*Mdn* $r = .78$). However, the reliabilities vary somewhat according to different age levels, with the younger subgroups having lower reliabilities than older groups.

Test-retest reliabilities are likewise quite high for the three IQ scores and somewhat lower for the specific subtests. Full Scale IQ reliability for all ages over a 23-day (median) retesting was .94 and the Verbal and Performance Scales were .94 and .87, respectively (Wechsler, 1991). The average increase in scores for retesting over the 23-day

interval was 7 to 8 points for the Full Scale IQ, 2 to 3 points for the Verbal IQ, and 11 to 13 points for the Performance IQ. This can mainly be accounted for by practice effects that seem to be particularly pronounced for the Performance Scale. The practical implication is that clinicians should incorporate the meaning of these short-term increases into their interpretations. Specifically, moderate short-term increases in scores of 5 to 10 points should not usually be considered to indicate true improvement in ability. Longer term retesting for the WISC-R over a two-year interval (which is more typical in clinical settings) has shown somewhat more stability with less than an average three-point difference in Full Scale IQ (Haynes & Howard, 1986). This suggests similar long-term test-retest stability for the WISC-III although no longer term studies are currently available. Test-retest reliabilities for the specific subtests ranged from a high of .89 for Vocabulary to a low of .69 for Object Assembly with an overall median of .76.

The standard error of measurement (indicated in IQ points) for the Full Scale IQ was 3.20, Verbal IQ was 3.53, and Performance IQ was 4.54. The standard error of measurement (given in subscale scores) for the Verbal subtests ranged from 1.08 to 1.45, with the narrowest range of error for Vocabulary (1.08) and the widest for Comprehension (1.45). The Performance subtests ranged from 1.11 to 1.67, with the narrowest range for Block Design (1.11) and widest for Object Assembly (1.67) and Mazes (1.64). Further information for incorporating specific standard error of measurement scores into WISC-III (and WAIS-III) interpretations is included in the Interpretation Procedures section.

The underlying factor structure, while still somewhat controversial, has generally supported Wechsler's conceptualization of abilities into a Verbal Comprehension factor that roughly corresponds with the Verbal Scale, and a Perceptual Organizational factor that generally corresponds with the Performance Scale (Kaufman, 1975, 1994; Sherman, Strauss, Slick, & Spellacy, 2000). More importantly, four factors have emerged from the WISC-III comprising Verbal Comprehension, Perceptual Organization, Freedom from Distractibility (Working Memory), and Processing Speed (Grice, Krohn, & Logerquist, 1999). This is comparable to the factors identified on the WAIS-III and also allows an attractive means of interpreting more specific aspects of intelligence than can be found using only IQ scores.

Given the high degree of item overlap, subtest correlations, and IQ score correlations between the WISC-R and WISC-III, much of the extensive validity research on the WISC-R can be generalized to the WISC-III (Dixon & Anderson, 1995). This validity relates primarily to extensive correlations with relevant criterion measures, including other ability tests, school grades, and achievement tests. Selected median correlations reviewed and reported by Sattler (2001) include those for the Stanford-Binet: Fourth Edition (.78), K-ABC (.70), group IQ tests (.66), WRAT (.52 to .59), Peabody Individual Achievement Test (.71), item overlap with the WPPSI-R (Sattler & Atkinson, 1993), and school grades (.39).

ASSETS AND LIMITATIONS

Since their initial publication, the Wechsler intelligence scales have been used in numerous research studies and have become widely used throughout the world. Thus,

they are familiar to both researchers and practitioners and also have a long and extensive history of continued evaluation. This enormous research base allows practitioners to make relatively accurate predictions regarding clients. Inconsistencies between an individual's performance and relevant research can also be noted, alerting the practitioner that he or she needs to develop and pursue further hypotheses. Furthermore, the subtests are relatively easy to administer, and the accompanying manuals provide clear instructions, concise tables, and excellent norms.

Norms for both the WAIS-III and WISC-III represent a further clear strength. The size is adequate and, for the most part, has corresponded to the demographics of the U.S. census. Cross-national use has been developed through research on how residents in other countries perform. Oversampling on the WAIS-III was done for African American and Hispanics as well as on a wide range of educational and ability levels to better understand how these groups perform. A further important feature is that the WAIS-III was co-normed with the Wechsler Memory Scale-III (WMS-III) and the Wechsler Individual Achievement Test (WIAT). This means that a high degree of confidence can be placed in comparing scores among these three different tests. Finally, the WAIS-III has extended its age range to include the performance for persons in the 74 to 89 range. This is an important feature given the increases in knowledge related to this age group along with the expanding number of persons over 65. One of the findings, for example, is that the Perceptual Organization and Processing Speed factors do not appear to be separate constructs for the 74 to 89 group.

Perhaps of even more practical importance to the clinician is the clear, precise data obtained regarding the person's cognitive functioning from the IQ, index, and subtest scores. For example, high scores on the Verbal Comprehension Index indicate good verbal abilities and that the person has benefited from formal education. In contrast, a low score on Processing Speed suggests the person would have a difficult time processing information quickly. A clinician can become extremely sensitive to the different nuances and implications of various patterns of scores. Thus, many of these interpretive guidelines, particularly for the IQ and index scores, have substantial theoretical and empirical support.

A final, but extremely important, asset of the Wechsler scales is their ability to aid in assessing personality variables. This can be done by directly observing the individual as he or she interacts with the examiner, studying the content of test item responses, or evaluating information inferred from the individual's pattern of subtest scores. For example, a person scoring low on Digit Span, Arithmetic, and Digit Symbol is likely to be experiencing anxiety, to have an attentional deficit, or a combination of both. On the other hand, it might be hypothesized that a person who scores high on both Comprehension and Picture Arrangement is likely to have good social judgment. Despite attempts to establish descriptions of the manner in which different clinical groups perform on the Wechsler intelligence scales, few clear findings have emerged (Piedmont, Sokolove, & Fleming, 1989a, 1989b). Thus, the Wechsler scales should not be seen as similar to "personality scales" or "clinical scales." Rather, the subject's subtest patterns, behavior surrounding the test, and qualitative responses to the items should be considered as a means of generating hypotheses related to personality. In this context, the Wechsler intelligence scales are noteworthy in the degree to which they can provide personality variables and clinical information.

One significant criticism leveled at the Wechsler scales has been their lack of data supporting their ecological (or everyday) validity (Groth-Marnat & Teal, 2000; Reinecke, Beebe, & Stein, 1999; Sbordone & Long, 1996). This is particularly important as referral questions are increasingly related to a client's everyday levels of functioning (i.e., extent of disability, ability to function independently, everyday aspects of memory). Although the Wechsler scales have been correlated with other measures, including the Stanford-Binet and academic achievement, for the most part, there has been a notable lack of comparisons with behavior external to the scales themselves. This is true despite the belief that many significant areas of a person, such as adaptive behavior, personal competence, or need for achievement, are separate (but related) constructs (Greenspan & Driscoll, 1997; Sternberg et al., 1995). In particular, the meanings associated with subtest scores should be investigated in more depth. For example, Picture Completion has traditionally been considered a measure of a person's ability to distinguish relevant from irrelevant details in his or her environment, yet this assumption has not been adequately tested. Likewise, no studies have been made to determine if high or low Digit Span scores relate to actual day-by-day behaviors, such as recalling telephone numbers, facility with computer programming sequences, or following directions.

An extension of this concern is that a number of authors have criticized what they believe is an overinterpretation of subtest and index scores (Glutting, McDermott, Konold, Snelbaker, & Watkins, 1998; Konold, Glutting, McDermott, Kush, & Watkins, 1999; MacMann & Barnett, 1997). Specifically, they believe that individual subtest reliabilities are too low and not sufficiently specific for interpreting individual profiles. For example, they note that, compared with *g* (as represented by the Full Scale IQ), WISC-III index scores do not account for a sufficient proportion of the variance in predicting achievement. As a result, index interpretation does not demonstrate sufficient incremental increases in prediction. In addition, the ipsative patterns of subtest strengths and weaknesses are not sufficiently stable over time (MacMann & Barnett, 1997). Clinicians might, therefore, be advised to rely on the Full Scale IQ rather than index scores when making academic (and possibly other) predictions. Various authors counter this by emphasizing the importance of hypothesis testing, combining interpretations with external criteria, and noting the conceptual importance of the complexity of intelligence (A. Kaufman, 1994, 1999; A. Kaufman & Lichtenberger, 2000, 2002; Lezak, 1988, 1995; Milberg et al., 1996).

There are several additional limitations to the Wechsler scales. Some critics believe that norms may not be applicable for ethnic minorities or persons from lower socioeconomic backgrounds. In addition, the complexity of scoring, particularly the numerous calculations required for the WAIS-III, is likely to increase the probability of clerical errors by examiners (Slate & Hunnicutt, 1988; Slate, Jones, & Murray, 1991). A further potential difficulty is that when supplementary subtests are substituted for regular subtests, it is unclear how these supplementary subtests will affect the three IQ or index scores. As a result, supplementary subtests should be given only under unusual circumstances, such as when one of the regular subtests has been "spoiled."

A further issue is that there is a certain degree of subjectivity when scoring many of the items on Comprehension, Similarities, and Vocabulary. Thus, a "hard" scorer may develop a somewhat lower score than an "easy" scorer. This is particularly true for

Similarities, Comprehension, and Vocabulary, where scoring criteria are less clear than for other subtests. The Wechsler scales, like other tests of intelligence, are also limited in the scope of what they can measure. They do not assess important factors, such as need for achievement, motivation, creativity, or success in dealing with people.

It should finally be noted that the WAIS-III and WISC-III have continued the traditional measurement of intelligence as represented by the Stanford-Binet scales and the earlier versions of the Wechsler scales. Although their revisions have provided features such as updated norms and index scores (especially the inclusion of Working Memory/Freedom from Distractibility and Processing Speed), the underlying theories and essential construction of these scales have remained relatively unchanged for well over 50 years. This is despite numerous developments in both theory and measurement. These include Luria's PASS (**P**lanning-**A**ttention-**S**uccessive-**S**equencing: Luria, 1980) model, Gardner's independent competencies (Gardner, 1999), various theories on emotional intelligence (Bar-On, 1998; Ciarochi, Chan, & Caputi, 2000), and commonsense problem solving (Sternberg et al., 1995). Thus, one criticism of the Wechsler intelligence scales is that they have not responded to more current views on intelligence (Kaufman & Lichtenberger, 2002; Sternberg, 1999; Sternberg & Kaufman, 1998; Styles, 1999). It remains to be seen whether newer models and assessment tools will have much of an impact on assessing either intelligence or, especially, the frequency to which the Wechsler scales will be used in this process.

MEANING OF IQ SCORES

Because only a weak and vague relation exists between theories of intelligence and the Wechsler intelligence scales, it is important for all persons involved with testing to understand the meaning of IQ scores. Untrained persons are particularly likely to misinterpret IQ scores, which may result in poor decisions or negative attitudes toward either the client or the testing procedure itself. The meaning of IQ scores can be partially clarified by elaborating on some of the more common misinterpretations. IQ is often incorrectly believed to be fixed, unchangeable, and innate. Although there does tend to be considerable stability of IQ scores throughout adulthood ($r = .85$; Schuerger & Witt, 1989) it is possible for changes in IQ to occur, particularly among children (Perkins & Grotzer, 1997). For example, the greatest longitudinal increases in IQs occurred among children who were from homes that provided strong encouragement and avoided severe forms of punishment (McCall, Appelbaum, & Hogarty, 1973). Similarly, Sameroff, Seifer, Baldwin, and Baldwin (1993) found that multiple environmental risk factors (e.g., number of major stressful events, mother's mental health) were able to predict one-third to one-half of IQ variance between the ages of 4 and 13. In addition, education can increase aspects of IQ primarily related to crystallized intelligence even among adults. Thus, IQ can be related to a number of environmental influences. Second, IQ scores are not exact, precise measurements; rather, they are estimates in which there is an expected range of fluctuation between one performance and the next. Finally, tests such as the Wechsler scales measure only a limited range of abilities, and a large number of variables usually considered "intelligent" are beyond the scope of most intelligence tests. No test or battery of tests can ever give a complete picture; they can only assess various areas of functioning. In summary, an IQ is

an estimate of a person's current level of functioning as measured by the various tasks required in a test.

An assumption of any global IQ score is that it derives from a wide array of interacting abilities. A subtest such as Information assesses specific areas of a person's range of knowledge and is related to general intelligence. IQ scores are also influenced by achievement orientation, curiosity, culture, and the person's interests. More general prerequisites are that the client must comprehend what has been requested, be motivated to do well, follow directions, provide a response, and understand English. Factors such as persistence and drive are also likely to influence any type of task presented to the person. The tasks included in IQ tests are those, based on judgments by psychometrists, most valued by Western society. In other words, they relate to and are predictive of relevant skills outside the testing situation. It is certainly possible to test a much wider range of areas (as in Guilford's Structure of Intelligence), but either these are not routinely done, or many potential measures may be of little relevance to predicting academic achievement or vocational performance.

Despite the many relevant areas measured by IQ tests, practitioners need to observe some humility when making predictions based on them. Many persons with quite high IQs achieve little or nothing. Having a high IQ is in no way a guarantee of success, but merely means that one important condition has been met. In contrast, persons with relatively low IQs have more severe limitations placed on them. As a result of their relatively narrower range of options, predictions regarding their behavior tend to be more accurate. However, it is possible that persons with average or below average WAIS-III/WISC-III IQs may have high levels of interpersonal, practical, or emotional "intelligence," which may help them compensate for lower levels of formal intelligence.

Regardless of the person's IQ range, clinicians should be clear regarding the likely band of error (standard error of measurement). It is often useful to include the standard error of measurement in a report. For example, the WAIS-III Full Scale IQ has an average standard error of measurement of 2.30 (Psychological Corporation, 1997). Thus, a particular IQ has a 95% chance of being within ± 5 IQ points of a person's obtained IQ. The WISC-III has a slightly higher average standard error of measurement of 3.20 for the Full Scale IQ, 3.53 for the Verbal IQ, and 4.54 for the Performance IQ (Wechsler, 1991). Error can also be the result of unforeseen events beyond the context of IQ tests. Although 50% to 75% of the variance of children's academic success is dependent on nonintellectual factors (persistence, personal adjustment, curiosity), most of a typical assessment is spent evaluating IQ. Some of these nonintellectual areas might be quite difficult to assess, and others might even be impossible to account for. For example, a student might unexpectedly develop an excellent relationship with a teacher, which significantly changes his or her attitude toward school, thereby stimulating his or her interest to passionately pursue a specific area. Thus, any meaning attached to an IQ score should acknowledge the possible effects of uncertainty both in the measurement itself and from the wider context of the person's life.

Another important aspect of IQ is the statistical meaning of the different scores. Binet originally conceptualized intelligence as the difference between a person's mental age and his or her chronological age. This was found to be inadequate and has been replaced by the use of the deviation IQ. The assumption behind the deviation IQ is that intelligence falls around a normal distribution (see Figure 5.1). The interpretation of an IQ score, then, is straightforward because it gives the relative position of a person

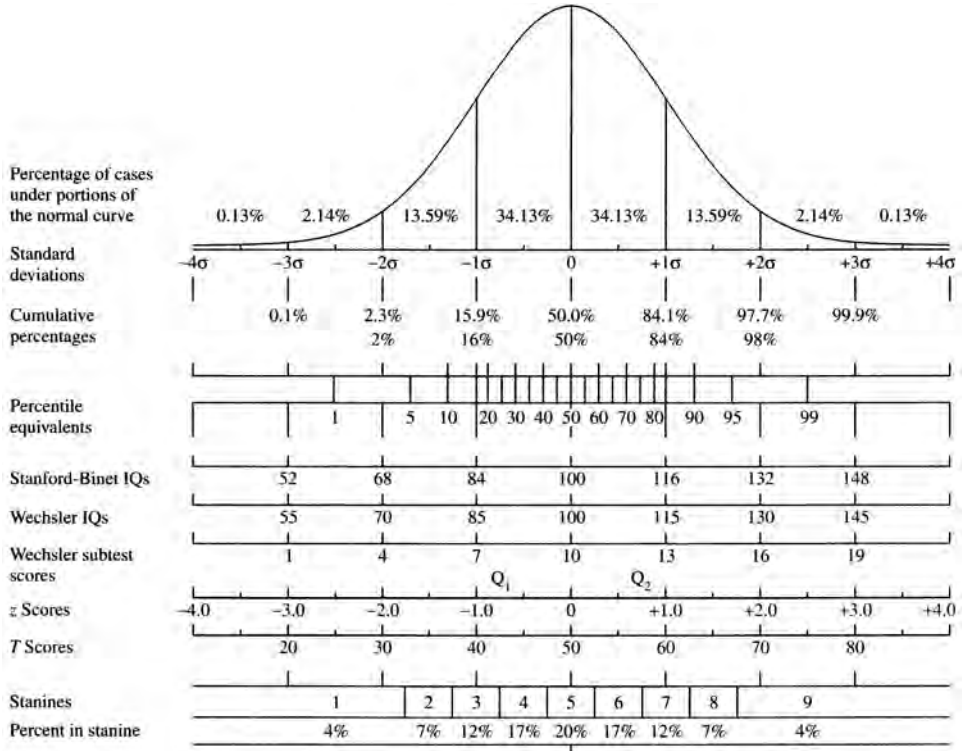

Percentage of cases under portions of the normal curve
0.13% 2.14% 13.59% 34.13% 34.13% 13.59% 2.14% 0.13%
Standard deviations
−4σ −3σ −2σ −1σ 0 +1σ +2σ +3σ +4σ
Cumulative percentages
0.1% 2.3% 15.9% 50.0% 84.1% 97.7% 99.9%
2% 16% 50% 84% 98%
Percentile equivalents
1 5 10 20 30 40 50 60 70 80 90 95 99
Stanford-Binet IQs
52 68 84 100 116 132 148
Wechsler IQs
55 70 85 100 115 130 145
Wechsler subtest scores
1 4 7 10 13 16 19
Q1 Q2
z Scores
−4.0 −3.0 −2.0 −1.0 0 +1.0 +2.0 +3.0 +4.0
T Scores
20 30 40 50 60 70 80
Stanines
1 2 3 4 5 6 7 8 9
Percent in stanine
4% 7% 12% 17% 20% 17% 12% 7% 4%

Figure 5.1 Relationship of Wechsler scores to various types of standard measures

compared with his or her age-related peers. The IQ can thus be expressed in deviation units away from the norm. Each of the three Wechsler IQs and four indexes has a mean of 100 and a standard deviation of 15. Scores also can be easily translated into percentile equivalents. For example, an IQ of 120 is 1.33 standard deviations above the mean and places an individual in the ninety-first percentile (see Appendix B on p. 677). Thus, this person's performance is better than 91% of his or her age-related peers. The IQ cutoff for mental retardation is 70, which indicates that such individuals are functioning in the lowest 2% when compared with their age-related peers. Appendix B can be used to convert Wechsler IQ scores ($M = 100$, $SD = 15$) into percentile rankings.

A final consideration is the different classifications of intelligence. Table 5.1 lists commonly used diagnostic labels and compares them with IQ ranges and percentages. These terms are taken from the 1997 WAIS-III manual. Thus, an IQ can be expressed conceptually as an estimate of a person's current level of ability, statistically as a deviation score that can be transformed into percentile equivalents, and diagnostically using common terms for classification.

CAUTIONS AND GUIDELINES IN ADMINISTRATION

The Wechsler manuals generally provide quite clear guidelines for administration and scoring. Despite this clarity, the number of administration and scoring errors on the

Table 5.1 Intelligence classifications

WAIS-III/WISC-III	More Value-Neutral Terms	Corresponding IQ Range
Very superior	Higher extreme above average	130+
Superior	Well above average	120–129
High average	High average	110–119
Average	Average	90–109
Low average	Low average	80–89
Borderline	Well below average	70–79
Extremely low (WAIS-III) or intellectually deficient (WISC-III)	Lower extreme	69 and below

Source: The classification systems of the WAIS-III are from Wechsler, 1997, Table 2.3, and for the WISC-III are from Wechsler, 1991, Table 2.8. Percentile ranks can be determined by consulting Appendix A.

part of trainee and experienced clinicians is far higher than they should be (Alfonso, Johnson, Patinella, & Radar, 1998; Moon, Blakey, Gorsuch, & Fantuzzo, 1991; Moon, Fantuzzo, & Gorsuch, 1986; Slate & Hunnicutt, 1988; Slate et al., 1991). Because the WAIS-III has a far greater number of calculations than the WAIS-R (or WISC-III), the likelihood (even probability) of clerical errors is significantly increased. One way of reducing clerical errors is to use the computer scoring software developed by The Psychological Corporation (i.e., Scoring Assistant for the Wechsler Scales for Adults, WAIS-III/WMS-III Writer, Scoring Assistant for the Wechsler Scales, WISC-III Writer). Even with repeated administration of the Wechsler scales, often examiners end up "practicing their mistakes" rather than correcting them (Slate et al., 1991). The causes of these errors include lack of proper instruction, lack of clarity between academic versus clinical site regarding where training is supposed to occur, carelessness, variations in the quality of the examiner-examinee relationship, and work overload for clinicians (Slate & Hunnicutt, 1988). One approach to reducing errors is awareness regarding the most frequent general categories of errors. These have been investigated by Slate et al. and the most common errors, in order of frequency, are as follows:

1. Failing to record examinee responses, circle scores, or record times (errors of administration).
2. Assigning too many points to an examinee's response (leniency by examiner).
3. Failing to question when required by test manual (poor reading and recalling of information in the manual).
4. Questioning examinee inappropriately (poor reading and/or incorrect integration of the manual).
5. Assigning too few points when required by test manual (examiner too hard).
6. Incorrectly converting raw score to standard score (clerical error).
7. Failing to assign correct points for Performance items (clerical and timing error).
8. Incorrectly calculating raw score for subtest totals (clerical error).
9. Incorrectly calculating chronological age (clerical error).

Whereas the preceding list covers quite general categories, the following list, adapted from Moon et al. (1991), includes a series of recommendations based on the most frequently occurring errors but does so by listing concrete and specific recommendations:

1. Recite digits (on Digit Span) and digits and letters (on Letter Number Sequencing) at the rate of one per second with the pitch of the voice dropping on the last digit/letter of each trial.
2. State during the introduction that each task begins with easy questions and ends with difficult ones. Examiners may also note that not everyone is expected to succeed on all problems.
3. Record responses verbatim on Vocabulary. At times, the examinee provides so much detail that this is not possible, but the essential components should be written down verbatim. This can be facilitated by the use of abbreviations.
4. Properly orient blocks (on Block Design) at examinee's midline.
5. The first time the examinee points out a nonessential part on Picture Completion, the examiner should comment, "Yes, but what is the most important thing missing?"
6. Attempt to elicit the examinee's perception of the testing situation and correct any misconceptions.
7. Check to see if the examinee is comfortable.

Despite clear guidelines in the manual as well as awareness of frequent errors, examiners are still likely to make mistakes. Thus, optimal training guidelines should be incorporated into graduate programs and continuing education. A recommended format is the Mastery Model, which involves the following steps: (a) 1 to 2 hours studying the manual, (b) viewing a videotape of a flawless WAIS-III/WISC-III administration, (c) viewing a videotaped lecture of major pitfalls of administration, (d) successfully detecting errors in a videotaped flawed WAIS-III/WISC-III administration, (e) actually administering the WAIS-III/WISC-IIIs to be evaluated by a rating device such as Sattler's (2001) "Administrative Checklist for the WAIS-III" (pp. 398–405) or "Administrative Checklist for the WISC-III" (pp. 243–248). Such procedures are likely to significantly shorten the length of training time, number of training administrations, and yet significantly increase the level of competence related to Wechsler scale administration and scoring (Moon et al., 1986; Slate et al., 1991).

The WAIS-III manual indicates that the average administration time to determine the Full Scale IQ for the standardization sample was 75 minutes (and 60 minutes for administering all subtests required to calculate only the indexes). In contrast, Ryan, Lopez, and Werth (1998) found that, for a heterogeneous clinical population, the average time was 91 minutes (and 77 minutes for subtests used to determine the indexes). Time estimates for the WISC-III standardization sample were 50 to 70 minutes (and the three supplementary subtests added an additional 10 to 15 minutes; Wechsler, 1991). These times were for administration only and did not include time required for scoring, breaks, or interpretation. The practical implications of this are that clinicians typically need to allocate more time for assessing clinical populations than might be inferred from reading the manual. Clients who fatigue easily may also need to have the

Wechsler intelligence scales administered over two sessions. Finally, clinicians should make realistic appraisals of required times and use these estimates to make sure that they are appropriately compensated.

WAIS-III/WISC-III SUCCESSIVE LEVEL INTERPRETATION PROCEDURE

The following successive-level approach to interpreting Wechsler scores represents an integration and synthesis of the approaches outlined by major resources in the field (A. Kaufman & Lichtenberger, 1999, 2000, 2002; J. H. Kramer, 1993; Naglieri, 1993; Sattler, 2001). This approach provides clinicians with a sequential, five-level format for working with and discussing a person's performance. The underlying purpose for each of these steps should be based on confirming, disconfirming, or altering hypotheses based on patterns of scores combined with relevant background information. The next section of this chapter ("Wechsler Subtests") covers descriptions of the Wechsler subtests, including the more frequently encountered abilities associated with these subtests. This section can serve as a summary and quick reference for clinicians, especially in analyzing test profiles (Levels II and III).

Examiners who are relatively unfamiliar with the Wechsler scales are likely to find the level of detail in the following interpretation procedure and Wechsler subtest sections somewhat daunting because of its complexity. It is thus recommended that they initially read the interpretation procedures to gain familiarity with the material. It might be particularly helpful to review the summary of these procedures in Table 5.2, both before and after reading this section. Table 5.2 can also serve as a useful future quick reference guide when actually working with Wechsler protocols. After perusing the "Interpretation Procedures" section, student examiners should next obtain a completed WAIS-III/WISC-III profile, preferably one they themselves have administered, and then work through the levels of interpretation in a sequential manner. This should add the required level of clarity and integration of the material to begin to work more confidently with future protocols.

The following are principles to keep in mind when working through the interpretation procedures:

- The successive steps begin with the most general aspects of the WAIS-III/WISC-III (Full Scale IQ) and gradually work their way to more specific aspects of the person's performance (indexes, additional groupings, subtest scatter, qualitative responses to individual items, etc.).
- Examiners can interpret the more global measures (Full Scale, Verbal, and Performance IQs) with greater meaning, usefulness, and certainty if there is not a high degree of subtest scatter. With increasing subtest scatter, the purity of the global measures becomes contaminated so that interpretations of these global measures become less meaningful. For example, if the Verbal scales display a pattern in which the Verbal Comprehension and Working Memory/Freedom from Distractibility Indexes are significantly different from each other, it makes more sense to focus on these two indexes rather than the more global Verbal IQ.

Table 5.2 Summary of successive five-level WAIS-III/WISC-III interpretive procedures

Level I. Interpret the Full Scale IQ
Determine percentile rankings and IQ classification (see Appendix B and Table 5.1)

Level II. Interpret Verbal-Performance, Factor Scores, and Additional Groupings

a. Verbal-Performance IQs.
Interpret if V-P discrepancy is significant (9 points for the WAIS-III and 12 points for the WISC-III).

b. Factor Scores: Verbal Comprehension, Perceptual Organization, Working Memory/Freedom from Distractibility, Processing Speed.
Interpret if significant discrepancies occur between the mean of the four WAIS-III/WISC-III factor scores.

c. Additional Groupings: Bannatyne's Categories, ACID/SCAD profiles, Horn groupings, Fuld profile (see Appendixes C and D).
Interpret if significant differences occur between means of groupings and individual grouping/category.

Level III. Interpret Subtest Variability (Profile Analysis)

a. Determine whether subtest fluctuations are significant:
   1. Decide appropriateness of using full scale versus verbal and/or performance subtest means; calculate relevant means.
   2. Calculate the difference scores between subtests and relevant means.
   3. Determine whether the difference between subtest score and scale means is significant (see Appendix E for WISC-III or "Score Conversion" page on WAIS-III Record Form).
   4. Indicate on profile as either a strength or a weakness.
   5. Repeat steps 1–5 for each relevant subtest.

b. Develop hypotheses related to the meaning of subtest fluctuations (Appendix F).

c. Integrate subtest hypotheses with additional information.

Level IV. Analyze Intrasubtest Variability.

Level V. Conduct a Qualitative Analysis.

- The recommended level set to establish significant difference is the .05 level. This is true for differences through all levels of interpretation including Verbal-Performance, indexes, additional groupings, and subtest differences. It was felt that this level of significance is sufficiently rigorous for clinical purposes. If either less stringent ($p = .15$) or more stringent ($p = .01$) levels are desired, relevant tables can be found in Wechsler (1991, 1997a) and Kaufman and Lichtenberger (1999, 2000, 2002). When possible, Bonferroni corrections have been included to correct for the possible statistical error resulting from inflation of significant results because of the number of comparisons.
- To determine whether index/factor scores are significantly (.05 level) discrepant, tables are consulted in the manuals (Table B.1, p. 205 in the *WAIS-III Administration and Scoring Manual* and Table B.1 in the *WISC-III Manual*). Thus, *comparisons are made between the different pairs of IQs/indexes/factors.*
- In contrast to the previous system, subtest fluctuations are based on *comparisons with mean scores.* One strategy is to compare the scaled score of each individual

subtest with the mean for all the subtests administered (and then calculate the difference that the subtests fluctuate from this mean to see if the difference is significant). A slightly different strategy is to compare each individual Verbal subtest with the overall mean for all Verbal subtests and also compare each individual Performance subtest with the overall mean for all Performance subtests. This latter method is appropriate if there is a significant difference between the Verbal and Performance IQs (9 points for the WAIS-III and 12 points for the WISC-III).

- Any interpretations, especially those related to the more specific levels (Levels III, IV, and V), should be considered as tentative hypotheses requiring support from additional sources of information (behavioral observations, school records, etc.). Preferably, each hypothesis should be supported by at least two additional sources. This process of hypothesis generation, confirmation/disconfirmation, and integration with other sources is not merely a statistical procedure but also involves considerable clinical wisdom and judgment.

Level I. The Full Scale IQ

An examinee's Full Scale IQ should be considered first because it provides the basis and context for evaluating other cognitive abilities. It is generally the single most reliable and valid score. The Full Scale IQ gives the person's relative standing in comparison with his or her age-related peers and provides a global estimate of his or her overall mental abilities. It is often useful to transform the Full Scale IQ into a percentile rank (see Appendix B) or intelligence classification (see Table 5.1). This is especially important when relating test results to untrained persons because both percentile rank and intelligence classifications are usually less subject to misinterpretation than IQ scores. Many examiners also prefer to include the standard error of measurement (SEM) as an estimate of the confidence that can be placed in the obtained score. For example, a WAIS-III Full Scale IQ of 110 has a 95% probability of falling between 105 and 115 IQ points. This clarifies that the IQ score is not a precise number but is rather a range with an expected margin of error. One classification, "Borderline," might potentially be misinterpreted, because it might be confused with the *DSM-IV* psychiatric diagnosis of *Borderline Personality.* Examiners might counter this by clarifying in parentheses that the "Borderline" range can also be described as "Well below Average."

Although the Full Scale IQ is the most stable and well-validated aspect of the Wechsler scales, its significance becomes progressively less important as the fluctuations increase between Verbal and Performance IQs, with high fluctuations between the factor scores, or when there is a high degree of subtest scatter. When such fluctuations occur, it is incumbent on the examiner to work in more detail to extract the significance of these relative strengths and weaknesses. The next four successive levels of interpretation provide a sequential method of accomplishing this goal.

Level II. Verbal-Performance IQs, Indexes, and Additional Grouping

The second level of interpretation is to consider Verbal and Performance IQs, index scores, and additional groupings. The .05 level of significance is consistently used to

determine if fluctuations are significant. In some cases, procedures and formulas are provided to determine the significance of various fluctuations and to convert scores into the familiar IQ-related standard scores having a mean of 100 and standard deviation of 1 (see summary of formulas in Appendixes C and D on pp. 678 and 679).

Step IIa. Verbal-Performance IQs

The Verbal IQ is an indication of an individual's verbal comprehensive abilities, while the Performance IQ provides an estimate of his or her perceptual organizational abilities. However, clinicians should be aware that a pure test of verbal comprehension or perceptual organization does not exist. A seemingly simple task, such as repeating numbers, involves not only verbal comprehension but also adequate rapport, ability to concentrate, number facility, and adequate short-term memory.

One of the central principles behind interpreting Verbal and Performance IQs is that there needs to be a significant difference between them. If such a difference occurs, an explanation for these differences should be developed. An area of difficulty (and controversy) lies in deciding what should be considered an interpretable difference. On the WAIS-III, a significant difference (at the .05 level) is 9 or more points. On the WISC-III, a 12-point difference is significant at the .05 level. It should still be noted that a full 18% of the WAIS-III and 24% of the WISC-III standardization samples obtained Verbal-Performance differences of 15 points or greater (Wechsler, 1981, 1991, 1997b; Psychological Corporation, 1997). This means that, although differences in the range of 9 to 15 points are statistically significant, they are still fairly common occurrences. The difference, then, may represent merely useful information rather than "lurking pathology." The possibility of pathology is far more likely with a 22-point discrepancy on the WAIS-III (occurred in only 5% of the standardization sample) or a 25-point discrepancy on the WISC-III (occurred in only 5% of the standardization sample).

Interpreting the magnitude of and meaning behind Verbal-Performance differences should always occur in the context of what is known about the person (particularly age and education) as well as his or her condition. For example, persons from higher socioeconomic backgrounds or with higher IQs are likely to have verbal scores significantly higher than their performance scores (R. A. Bornstein, Suga, & Prifitera, 1987; Wechsler, 1997a). In contrast, unskilled workers are more likely to have higher performance scores relative to verbal. If these trends are reversed (e.g., an attorney with higher performance scores or unskilled worker with higher verbal than performance scores), the importance of such a result becomes greater. The major variables influencing, and possible meanings behind, Verbal-Performance score differences are summarized in the sections "Verbal Scales" and "Performance Scales," as well as in the sections on special populations. However, Sattler (2001) has summarized the possible general meanings associated with such differences as relating to cognitive style, patterns of interests, sensory deficits, psychopathology (such as emotional disturbance or brain damage), deficiencies/strengths in information processing, or deficiencies/strengths in ability to work under pressure. After an interpretation has been made, practitioners can eventually work to develop implications and instructional recommendations for high and low scores (for Full Scale, Verbal, and Performance IQs, and Working Memory/Freedom from Distractibility).

Under certain conditions, even statistically significant differences between Verbal and Performance IQs can be considered meaningless. The first condition occurs when the Verbal scale splits into significant differences (10 points or more for the WAIS-III; 13 points or more for the WISC-III) between the Verbal Comprehension and Working Memory/Freedom from Distractibility Indexes. This separation means that the Verbal scales do not cohere into a unitary construct. Thus, it may make more sense to interpret the two index scores separately (see interpretative level IIb). Similarly, if a significant difference (13 points for the WAIS-III; 15 points for the WISC-III) exists between Perceptual Organization and Processing Speed, it means that the Performance IQ is not coherent. Accordingly, the two indexes should be interpreted separately. A further condition that might render the Verbal and/or Performance IQs meaningless occurs when there is a high degree of subtest scatter in general (WAIS-III Verbal subtest range 8+ points, Performance 8+ points; WISC-III Verbal range 7+ points, Performance range 9+ points). This happens because the intent of the three IQs represents a unitary construct in which the person's Full, Verbal, or Performance IQs are general, integrated means of functioning. In contrast, high subtest scatter attacks the unitary, integrated nature of the IQs. It is then the examiner's task to work with the relative high and low combinations of subtests to make sense of the person's intellectual strengths and weaknesses. These steps are outlined in Levels IIb, IIc, and III. However, before continuing to an interpretation of subtest scatter, important clusters of subtests might be found through a consideration of indexes and additional groupings.

Step IIb. Index Scores

The WAIS-III and WISC-III have both been found to have four different index scores. Their interpretation provides an empirically and conceptually based means of understanding more detailed aspects of the person's intellectual functioning (see interpretive summaries in Appendix E). As with the Verbal and Performance Scales, the indexes should be interpreted only if discrepancies between the highest and lowest subtests comprising the indexes are not too large (WAIS-III: Verbal Comprehension, 5 points; Perceptual Organization, 6 points; Working Memory, 6 points; Processing Speed, 4 points. WISC-III: Verbal Comprehension, 7; Perceptual Organization, 8; Freedom from Distractibility, 4; Processing Speed 4).

The Verbal Comprehension Index (WAIS-III: Vocabulary, Similarities, and Information; WISC-III: Information, Similarities, Vocabulary, and Comprehension) is a purer, more refined measure of verbal abilities than the Verbal Scale itself. This has been achieved by excluding Digit Span and Arithmetic (and Letter-Number Sequencing on the WAIS-III), which focus primarily on sequential processing, number ability, attention, and working memory rather than strictly verbal abilities. The material presented to them is in the form of oral questions that they need to answer. As such, an examinee's score on Verbal Comprehension reflects the extent to which they understand the meanings of words, can conceptualize verbal information, the extent of factual knowledge related to verbal material, and their ability to adequately express the material in words.

The Perceptual Organization Index (POI) is likewise a somewhat purer measure of perceptual abilities (WAIS-III: Picture Completion, Block Design, Matrix Reasoning; WISC-III: Picture Completion, Picture Arrangement, Block Design, and Object

Assembly). Perceptual Organization is less a measure of processing speed than the Performance Scale because only one subtest (Block Design) emphasizes speed. An examinee's score reflects the extent to which he or she has good nonverbal, fluid reasoning; can integrate nonverbal material; pays close attention to detail; and accurately responds to the visual-spatial material presented to him or her. Much of this involves using the kind of visual-spatial and visual-motor skills to solve problems that are not taught in formal academic schooling.

Working Memory/Freedom from Distractibility (WAIS-III: Arithmetic, Digit Span, Letter-Numbering Sequencing; WISC-III: Arithmetic and Digit Span) is a more complex and controversial construct and has been extensively studied with children but much less so with adults. It has primarily been related to concentration, attention, and short-term memory and is believed to be lowered by poor number facility, anxiety, difficulty making mental shifts, and poor self-monitoring. Sequencing is also crucial because each of the relevant subtests requires that the respondent place numbers and symbols in their proper order. Wielkiewicz (1990) has suggested that the low concentration, memory, and sequencing reflected on this factor is often because of a poorly functioning executive ability. Specifically, the person experiences difficulty attending to stimuli and simultaneously performing other mental tasks (e.g., listening to spoken digits and storing them while simultaneously reversing them and then repeating them backward). Good performance also requires a high level of motivation. As a result of these diverse functions, a low Working Memory/Freedom from Distractibility factor is also likely to lower performances in other areas, and this should be considered when estimating the person's overall potential.

It is crucial to consider a variety of interpretive possibilities to interpret the Working Memory/Freedom from Distractibility Index. Often behavioral observations can be crucial. A client who frequently asks to have the questions repeated might have a high level of distractibility. Alternatively, a high degree of motor activity or excessive talking might highlight a client's high level of anxiety. If number skills have not been developed, the client might ask to write out the numbers related to the arithmetic problems or count out the numbers with his or her fingers. The importance of cautiously interpreting (and providing additional support) for Freedom from Distractibility is highlighted because Reinecke et al. (1999) were unable to find an association between Freedom from Distractibility and children diagnosed with ADHD.

The Processing Speed Index (PSI; WAIS-III/WISC-III: Symbol Search and Coding) reflects the mental and motor speed with which a person can solve nonverbal problems. Further subtest support for this index can be found if the person also has correspondingly high (or low) performances on the timed nonverbal tests of Object Assembly and Block Design. In addition to mental and motor speed, the Processing Speed factor is also a measure of a person's ability to plan, organize, and develop relevant strategies. Low scores on Processing Speed can also reflect poor motor control. Because speed and concentration require good test-taking attitudes, Processing Speed (as well as Freedom from Distractibility) can also be lowered by poor motivation to perform well. For this reason, these two factors are sometimes referred to as *validity* factors. Whether a lowered performance is the result of poor motivation can often best be assessed by behavioral observations in combination with clarification and consideration of the presenting problem. An overly reflective problem-solving style could also

lower the Processing Speed factor because the person would take too much time cautiously considering his or her response to each item.

Both the WAIS-III and WISC-III have made calculating index scores easy; that is, conversion into standard scores ($M = 100$, $SD = 15$) has been incorporated into the normal scoring procedure. However, examiners still need to determine the significant (interpretable) differences between the combinations of index scores. This can be done for the WAIS-III by consulting Table B.1 in the *WAIS-III Administration and Scoring Manual* (Wechsler, 1997a, p. 205) and summarizing the results on the "Discrepancy Comparisons" table on the Discrepancy Analysis Page. The significance of the WISC-III Index score discrepancies can be determined by consulting Table B.1 in the *WISC-III Manual* (Wechsler, 1991, p. 261). There is no table to summarize the WISC-III index discrepancies on the WISC-III Record Form. Interpretations are then based on the comparisons between the different indexes. For example, if a person's WAIS-III Perceptual Organization Index was 15 points higher than his or her Processing Speed Index (POI > PSI), this difference would clearly exceed the .05 level and, according to Table B.2, a POI-PSI difference of 15 points occurred in only 32.4% of the standardization sample. It should be noted that this figure combines both POI > PSI and PSI > POI. Thus, it would be more accurate to half the percentage ($32.4 \div 2 = 16.2\%$) because there is interest only in knowing the frequency which POI > PSI occurs (and not PSI > POI). This results in a rounded-off frequency of 16%.

Index scores can be used for interpreting a person's relative strengths and weaknesses. However, the actual factor-based standard scores should not be presented in the psychological report because readers may confuse them with IQ scores. In addition, including both IQ scores and factor-based standard scores would make the report unnecessarily cluttered with too many numbers. After interpretations have been made, practitioners can then work to develop appropriate instructional recommendations if an educational or rehabilitation plan needs to be developed.

Step IIc. Additional Groupings: Bannatyne's Categories, ACID/SCAD Profiles, Horn Groupings

Four additional factors or groupings can often yield useful interpretations. These are optional and should be calculated when, on initially appraising the pattern of subtest scatter, it seems they might be relevant to investigate more formally. For example, if subtests that are highly loaded on spatial abilities (Picture Completion, Block Design, Matrix Reasoning, Object Assembly) appear significantly higher than sequencing subtests (Digit Span, Arithmetic, Digit Symbol), a formal calculation of Bannatyne's categories will serve to confirm or disconfirm initial impressions related to the subtest profiles. Another reason to calculate the groupings listed in Level IIc occurs when an examiner wishes to see if a person's subtest profile is similar or dissimilar to a person from an actual or suspected client population (e.g., learning disabled, Alzheimer's disease).

Bannatyne's Categories Bannatyne's categories comprise subtest patterns in which *Spatial* abilities are relatively higher than *Verbal Conceptualization,* which is in turn higher than *Sequential* abilities, with *Acquired Knowledge* typically being the lowest (Verbal Conceptualization > Spatial > Sequential > Acquired Knowledge).

These categories were originally developed as a means of detecting and understanding learning-disabled populations (Bannatyne, 1974). However, it has been found that many learning-disabled persons do not necessarily have this pattern, and many non-learning-disabled populations do have the pattern (see Groth-Marnat, 2002 and subsection "Learning Disabilities"). The result has been that Bannatyne's categories have been used to further understand Wechsler scale profiles in general and not for the diagnosis of specific conditions.

The recommended method of interpreting the WAIS-III Bannatyne's factors is to first use the following formulas to transform the subtest groupings comprising each of the factors into the familiar standard scores (Mean of 100 and Standard Deviation of 15). Note that scaled scores must be used.

WAIS-III Verbal Conceptualization: 1.8 (Vocabulary + Similarities + Comprehension) + 46

WAIS-III Spatial: 1.5 (Matrix Reasoning + Block Design + Object Assembly + Picture Completion) + 40

WAIS-III Sequential: 1.6 (Arithmetic + Digit Span + Digit Symbol + Letter-Number Sequencing) + 36

WAIS-III Acquired Knowledge: 1.9 (Information + Arithmetic + Vocabulary) + 43

An appraisal of the standard scores gives a general idea as to whether the classic Bannatyne pattern is present. However, to more precisely determine whether or not the differences are significant (and therefore interpretable), the following additional procedures must be taken. First, find the means of the standard scores and then subtract the mean from each of the standard scores. This indicates the various differences between the mean and the standard scores. To be significant at the .05 level, the following difference scores must be equal to or greater than:

Verbal Conceptualization	7.1
Spatial	7.9
Sequential	7.5
Acquired Knowledge	6.6

Note that the inclusion of the WAIS-III Matrix Reasoning and Letter-Number Sequencing subtests is somewhat speculative at this time but their inclusion is conceptually consistent because these subtests are measuring spatial and sequencing abilities, respectively. Picture Arrangement has been excluded from the categories because it is a maverick subtest that doesn't clearly load on any of the four categories.

The procedure for interpreting Bannatyne's factors on the WISC-III is similar. First, the subtest scores comprising each of the groupings should be transformed into standard scores using the following formulas:

WISC-III Verbal Conceptualization: 1.9 (Vocabulary + Comprehension + Similarities) + 43

WISC-III Spatial: 2.0 (Picture Completion + Block Design + Object Assembly) + 40

WISC-III Sequential: 2.3 (Digit Span + Arithmetic + Coding) + 31

WISC-III Acquired Knowledge: 1.9 (Information + Vocabulary + Arithmetic) + 43

As with the WAIS-III, the mean of the standard scores must then be calculated and the differences between the means must then be determined. To be significant at the .05 level, the following differences must be equal to or greater than the following values:

Spatial	10.0
Verbal Conceptualization	8.5
Sequential	12.0
Acquired Knowledge	8.5

Several interpretive warnings are advisable. Only about 20% to 25% of learning-disabled persons will actually demonstrate the classic sequence of Bannatyne's Verbal Conceptualization > Spatial > Sequential > Acquired Knowledge pattern (see Groth-Marnat, 2002 and A. Kaufman & Lichtenberger, 2002). One noteworthy variation from the Bannatyne profile is that sometimes a bright, highly motivated learning-disabled person with poor sequencing abilities compensates by developing a high level of acquired knowledge. Thus, the Acquired Knowledge category might be outstandingly high even though Sequential abilities might still be quite low. Another less bright and/or less motivated learning-disabled person might experience the disruption of poor sequencing to a greater extent and may have then become correspondingly alienated from academic learning. This would then be reflected in an outstandingly low Acquired Knowledge category. This is consistent with the finding that learning disabilities are a heterogeneous group of disorders with sometimes well-defined subtypes (A. Kaufman & Kaufman, 2002). This means that examiners need to take a flexible approach toward interpreting the relation and implications between the Bannatyne categories.

The ACID/SCALD/SCAD Profiles The ACID, SCALD, and SCAD profiles are similar to those of Bannatyne's categories. Low scores on each of these profiles have been found to occur more frequently among learning-disabled populations (Cordoni, O'Donnell, Ramaniah, Kurtz, & Rosenshein, 1981; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Wechlser, 1997b). The WAIS-III/WISC-III *ACID* profile comprises **A**rithmetic, **C**oding (Digit Symbol-Coding for the WAIS-III), **I**nformation, and **D**igit Span. Note that three of the subtests (Arithmetic, Coding/Digit Symbol, Digit Span) comprise Bannatyne's Sequential Category and one (Information) is included in Bannatyne's Acquired Knowledge category. As with the Bannatyne categories, an exception to the pattern is that often learning-disabled university students who have academically compensated for their learning disabilities have relatively good performances on Information (Ackerman et al., 1987). A new *SCALD* profile has been proposed for the WAIS-III (A. Kaufman & Lichtenberger, 1999, 2002) and comprises **S**ymbol Search, **C**oding (Digit Symbol-Coding), **A**rithmetic, **L**etter-Number Sequencing, and **D**igit Symbol. This

is based on the finding that the Working Memory and Processing Speed indexes were generally lowest among adult learning disabled persons (Psychological Corporation, 1997) and are composed of these subtests.

For the WISC-III, A. Kaufman (1994) recommends deleting Information and instead inserting Symbol Search. The profile then comprises **S**ymbol Search, **C**oding, **A**rithmetic, and **D**igit Span and can then be appropriately renamed the *SCAD* profile. These four subtests are a merging of the WISC-III's Freedom from Distractibility (Arithmetic and Digit Span) and Perceptual Speed (Symbol Search and Coding) factors. To convert the SCAD profile into a standard score, the following formula can be used (A. Kaufman, 1994):

$$\text{SCAD (WISC-III): } 1.7\ (\text{SS} + \text{C} + \text{A DSp}) + 32$$

To vary from the WISC-III Full Scale IQ (standard score for all subtests) at the .05 level, the standard SCAD score must be 9.5 points above/below the Full Scale IQ. However, if there is a significant difference (16 or more points) between the Processing Speed (Symbol Search and Coding) and Freedom from Distractibility (Arithmetic and Digit Span) indexes, the SCAD profile should not be interpreted.

Horn Groupings Horn and Cattell's (1966) fluid versus crystallized intelligence has been used to organize many of the Wechsler intelligence scales (J. Caruso & Cliff, 1999; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Woodcock, 1990). On the WAIS-III, *Fluid Intelligence (Gf)* primarily includes Digit Span and Matrix Reasoning whereas *Crystallized Intelligence (Gc)* is measured by Vocabulary and Information (J. Caruso & Cliff, 1999). Kaufman and Lichtenberger (1999, 2002) have conceptually reorganized the WAIS-III subtests around the more detailed Horn (1985) groupings of *broad visualization* (*Gv*; Matrix Reasoning, Block Design, Object Assembly, Picture Completion), *broad speediness* (*Gs;* Digit Symbol-Coding, Symbol Search, Object Assembly) and *short-term memory* (Letter-Number Sequencing, Arithmetic, Digit Span). A rough idea of the examinee's relative strengths and weaknesses can be obtained by adding the subtest scaled scores, finding the mean, and comparing the strength and weaknesses. A more precise method is to convert the groupings into standard scores using the following formulas (A. Kaufman & Lichtenberger, 2002):

$$\text{WAIS-III Fluid Intelligence: } 1.1\ (\text{MR} + \text{BD} + \text{OA} + \text{S} + \text{PA} + \text{A}) + 34$$

$$\text{WAIS-III Crystallized Intelligence: } 1.2\ (\text{I} + \text{V} + \text{C} + \text{S} + \text{PA}) + 40$$

$$\text{WAIS-III Broad Visualization: } 1.5\ (\text{MR} + \text{BD} + \text{OA} + \text{PC})\ 40$$

$$\text{WAIS-III Broad Speediness: } 1.9\ (\text{CD} + \text{SS} + \text{OA}) + 43)$$

WAIS-III Short-Term Memory: (Same as Working Memory Index)

After these standard scores have been determined, examinees next should find the mean of the five standard scores and calculate the differences that each of the groupings varies from the mean. To be significant at the .05 level (and thus interpretable), the following values should be either equal to or greater than the following:

Fluid Intelligence	7.5
Crystallized Intelligence	7.0
Broad Speediness	10.5
Short-Term Memory	8.3

The Horn groupings on the WISC-III are somewhat different from the WAIS-III in that Fluid Intelligence includes Picture Arrangement, Block Design, Object Assembly, Similarities, and Arithmetic. Crystallized Intelligence includes Information, Similarities, Vocabulary, Comprehension, and Picture Arrangement. Because Picture Arrangement and Similarities include skills related to both crystallized and fluid intelligence, they are included in both groupings. An additional grouping is Achievement, which is a composite of all tests most influenced by academic learning. This grouping includes Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Picture Arrangement. To convert the WISC-III Horn groupings into standard scores ($M = 100$, $SD = 15$), the following formulas can be used (A. Kaufman, 1994):

WISC-III Fluid Intelligence: 1.3 (S + A + PA + BD + OA) + 35

WISC-III Crystallized Intelligence: 1.3 (I + S + V + C + PA) + 35

WISC-III Achievement: 0.85 (I + S + A + V + C + PA) + 49

Comparisons should be made between each of the WISC-III three standard scores to determine whether they are significantly different. To do this, first calculate the mean for the total subtests used in the three Horn groupings by summing the three standard scores and dividing by three. Then calculate the difference that each of the three standard scores varies from the mean. To be significant at the .05 level, the following values must be achieved:

Fluid Intelligence: 8.5 points
Crystallized Intelligence: 9 points
Achievement: 8.5 points

Level III. Interpreting Subtest Variability

The third step is to consider the degree to which the individual subtests deviate from the full scale, verbal, or performance subtest means and to determine the meaning associated with the subtest fluctuations. The outcome should be a description of a person's relative cognitive strengths and weaknesses. A listing and discussion of the meaning of each subtest and the abilities it measures is provided in the next major section of this chapter ("Wechsler Subtests"). Clinicians can refer to this section, as well as to information on how to assess special populations in developing their own hypotheses about important dimensions of intersubtest scatter. Readers may also wish to refer to R. Gregory (1999), A. Kaufman (1990, 1994), A. Kaufman and Lichtenberger (1999, 2000, 2002), Naglieri (1993), and Sattler (2001), who have provided detailed lists of hypotheses and useful tables for various combinations of high and low subtest scores. However,

Level III interpretation is necessary only if there is sufficient subtest scatter. If all the subtests are fairly even, it is not necessary to attempt subtest profile interpretation.

Clinicians need to be aware that interpreting subtest variability involves clinical judgment guided by theory, observation, and an integration of the specifics of each case. Because there is little research base to support this process, it should be approached with caution. As a rule, the more subtests that can be combined to make inferences based on their shared abilities, the more support can be found for such an inference. At one extreme would be a series of subtests that combine to make up one of the previously discussed indexes/factors. The opposite would be only one subtest used to make an inference. In general, inferences based on only a single subtest should be treated with the most caution. While these single subtest-based inferences can be viable, it is incumbent on the clinician to obtain as much supporting evidence as possible.

It should also be noted that subtest interpretation has been the source of controversy in that some authors have pointed out that the subtests are not sufficiently reliable, do not have enough subtest specificity, and do not provide sufficient incremental validity beyond what might be accounted for by the Full Scale IQ (Konold et al., 1999; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992). In part, this relates to empirical concerns, but there are also underlying conceptual differences centered around whether intelligence is mainly accounted for by *g* ("lumpers") as opposed to its being composed of a number of different components ("splitters"). This debate seems to have been present almost as long as conceptions of intelligence have been in existence. One common response to this issue is that subtest interpretation is not merely an empirical activity but also involves a clinical process of hypothesis testing and integrating a variety of sources of data (see A. Kaufman, 1994; A. Kaufman & Lichtenberger, 1999, 2000, 2002; Lezak, 1995; E. Kaplan et al., 1999). Accordingly, the following three steps are recommended in interpreting subtest variability: (a) Determine whether subtest fluctuations are significant, (b) develop hypotheses related to the meaning of any subtest fluctuations, and (c) integrate these hypotheses with additional relevant information regarding the examinee. Clinicians should very clearly *not* attempt to interpret subtests by merely listing the abilities provided in the subtest descriptions. It is hoped that the following guidelines will help ensure that clinicians develop accurate, useful, and well-integrated interpretations of subtests.

Step IIIa. Determine Whether Subtest Fluctuations Are Significant

The first step in profile analysis is to account for the implications of the mostly modest reliabilities associated with the different subtests. This means each clinician needs to seriously consider whether the variability results from reliably measured strengths or weaknesses or is merely from chance error inherent in the subtest. The WAIS-III Record Form conveniently allows examinees to determine whether subtests are significantly different from the mean of all subtests included under the listing for Verbal subtests and from the mean of all subtests included under the listing for Performance subtests (see WAIS-III Score Conversion Page). This may be somewhat confusing because there are some subtests included (Symbol Search, Letter-Number Sequencing, Object Assembly) in calculating the "Verbal" and "Performance" means that are not actually used to calculate the Verbal and Performance IQs. However, they are still listed as verbal or performance subtests. Whether the discrepancies are significant can

be determined by calculating the magnitude that each WAIS-III subtest varies from the "Verbal" and "Performance" mean and noting whether it is significant based on Table B.3 in the *WAIS-III Administration and Scoring Manual* (pp. 208–209). The various subtests can then be indicated as either strengths or weaknesses ("Ss" or "Ws"). The WAIS-III also provides procedures for establishing the frequency with which the subtests fluctuate in the standardization population.

One area of potential confusion is that the WAIS-III manual does not provide tables to develop scatter analysis when six Performance Scale subtests are administered. Unfortunately, using six subtests is necessary if examiners would like to obtain both the Performance IQ and Processing Speed index score. There are also no tables available if certain optional subtests are used instead of the standard subtests. LoBello, Thompson, and Venugopala (1998) have provided supplementary tables to help examiners work with these situations.

The WISC-III does not provide a worksheet for determining subtest fluctuations. However, the magnitude (and whether the subtests are strengths or weaknesses) can be determined by following the guidelines detailed in Appendix E (Worksheet for Determining Magnitude of WISC-III Subtest Fluctuations). In most instances, the Worksheet for Determining Magnitude of WISC-III Subtest Fluctuations is sufficient for calculating whether subtests are relative strengths or weaknesses. If an unusual number of subtests have been administered, it might be necessary to consult Table B.3 in the *WISC-III Manual* (p. 263) to determine means and magnitudes of discrepancies. These guidelines will help clinicians determine whether the subtests actually fluctuate from the Verbal, Performance, or Full Scale means to a significant (at the .05 level) extent. It should be noted that a moderately high range (highest minus lowest subtest score) is a common occurrence. The average range on the WISC-III standardization sample was 8.5 points (*SD* = 2.3) for the Full Scale (13 subtests), 5.5 (*SD* = 2.1) points for the Verbal Scale (6 subtests), and 7.1 points (*SD* = 2.4) for the Performance Scale (7 subtests; Wechsler, 1991). Approximately two thirds of the WISC-III standardization sample had subscales that ranged between 7 and 8 points. Thus, clinicians should be cautious about not inferring pathology when such differences might merely indicate preferences for differing cognitive styles.

Often, it is found that there are no significant subtest fluctuations. In these cases, do not proceed to Steps IIIb and IIIc. Instead, focus on interpreting the profile based on information derived from Levels I and II and possibly Levels IV and V if these are relevant.

Step IIIb. Develop Hypotheses Related to the Meaning of Subtest Fluctuations

Just because a subtest or group of subtests has been designated as a relative strength or weakness does not mean that it is clear which of the various functions involved with the subtest is a strength or weakness. For example, Picture Arrangement involves planning (sequencing), visual organization, distinguishing essential from nonessential detail, and comprehending social situations. For one person, scoring high in Picture Arrangement might reflect excellent planning/sequencing abilities, for another it might reflect good social skills, and a third person might be high in both. It is the examiner's responsibility to become actively engaged with the pattern of subtests and any other

relevant sources of information to determine which ability or abilities are high and low for the person. Interpreters who merely list the subtest's abilities as they are listed in a book are quite likely to make incorrect and even potentially damaging conclusions about the examinee. This cookbook type of approach should be strongly discouraged.

The underlying principle in uncovering the actual subtest strengths or weaknesses is to initially consider a significantly high or low subtest score in the context of scores on other subtests. If a person has scored high on Picture Arrangement and this might reflect good planning/sequencing, an examiner would expect other subtests that also measure planning/sequencing to be at least within the average range, if not higher. Thus, the examiner might make sure that other sequencing-oriented tasks, primarily Digit Span, Arithmetic, and Letter-Number Sequencing, were also high.

The difficulty with such a procedure is that it requires an in-depth knowledge of each subtest's abilities, familiarity with frequent clusters of subtests, and an overreliance on intuition in terms of noticing and testing different patterns. This is a particularly daunting task for beginning and even experienced clinicians. Thus, a formal step-by-step procedure of comparing and contrasting relative strengths and weaknesses is recommended. This can be accomplished by completing Appendix F ("Guidelines for Hypothesizing Subtest Strengths and Weaknesses" p. 682). These guidelines use the same underlying principle in that consistencies and inconsistencies among patterns of subtests are determined. However, these patterns are investigated in a thorough and systematic pattern. The directions are adapted from A. Kaufman (1990, 1994) and A. Kaufman and Lichtenberger (1999, 2000, 2002) and the listed subtest abilities were adapted from those described by a wide variety of sources, including Bannatyne (1974), Horn (1985), A. Kaufman (1990, 1994), A. Kaufman and Lichtenberger (1999, 2000, 2002), Lezak (1995), and Sattler (2001). After completing Appendix F, the clinician will have arrived at a series of empirically derived and partially tested hypotheses.

An important consideration in this strategy of subtest interpretation is that it should not be a rigid, mechanical process. For example, a client who presents with subjective complaints related to poor sequencing (e.g., difficulty following directions, placing things in the wrong order) may not necessarily have all the expected WAIS-III/WISC-III subtests quite within the statistically interpretable range. However, given the quite clear symptom reports (and possibly behavioral observations), practitioners may still choose to interpret the sequencing-related subtests. In contrast, another client might have most sequencing subtests in the statistically significant range but poor sequencing was neither a symptom complaint, nor were behavioral observations noted that would have been consistent with poor sequencing. As a result, the hypothesis of poor sequencing might be rejected as not applying to the person. The outlined procedure, then, should be used for hypothesis generation in which other factors beyond the mechanical interpretation procedure can confirm or disconfirm these hypotheses.

Step IIIc. Integrate Subtest Hypotheses with Additional Information

Before finally accepting or rejecting the step-by-step empirically derived hypotheses from Steps IIIa and IIIb, examiners should consider additional sources of relevant information. This might include behavioral observations, medical records, school records, teacher's reports, other test data, or qualitative responses that examinees have made to the test items (see Level V). For example, an examiner might be trying to decide

whether low scores on Arithmetic and Digit Span reflect poor attention or poor sequencing. If the examinee was observed to have attended well to the tasks but had difficulty following a series of directions, then it suggests sequencing is more likely to be the difficulty. Or, an examiner might be trying to decide whether the examinee prefers a simultaneous or sequential style of processing information. A relevant behavioral observation is careful observation of the way the person worked on Block Design. Did he or she proceed in a step-by-step sequence, trying to match each block with a segment of the picture, or, rather, did he or she try to understand the design as a whole while attempting to complete the task? A final relevant example might be low scores on Arithmetic, Digit Span, Digit Symbol-Coding, and Symbol Search. Each of these subtests requires a high level of motivation. Indeed, they have sometimes been referred to as validity scales because they are likely to be lowered as a result of poor motivation (A. Kaufman, 1994). Rather than work to decipher the examinee's low abilities as reflected in these subtests, the examiner might decide that behavioral observations more accurately suggest the person was not expending a sufficient amount of effort.

A focus on additional sources of information, particularly behavioral observations, also has relevance for determining the significance of subtest fluctuations (Step IIIa) and developing hypotheses (Step IIIb). As was stressed previously, sometimes a subtest fluctuation may not quite achieve formal statistical significance, yet, because of additional information, the practitioner feels justified in giving the score greater clinical importance and considering it for interpretation. Similarly, generating hypotheses by formally putting the data through Step IIIb may not have confirmed a suspected hypothesis. However, if a clinician has additional information that might justify accepting the suspected hypothesis, he or she may be persuaded to accept it although some of the formal procedures have not quite supported it. This highlights an essential underlying philosophy of Wechsler scale and subtest interpretation: It is not solely a statistical and empirical exercise, but, more importantly, it involves the use of clinical skills and judgment.

Level IV. Intrasubtest Variability

A further, potentially important area of analysis involves looking at the patterns of performance within the items of each subtest. These items are arranged in sequences that become progressively more difficult. Thus, a normal and expected pattern would have the examinee pass the initial items and slowly but evenly begin to fail more difficult ones. A more sporadic pattern, in which the examinee misses initial easier items but passes later more difficult ones, may suggest an attentional deficit or specific memory losses, particularly related to retrieval difficulties (E. Kaplan, Fein, Morris, & Delis, 1991; E. Kaplan et al., 1999). If performance is highly sporadic, the reason should be explored further. For example, clients might be consciously faking if they miss every other item, miss extremely easy items, and/or appear much more alert than their obtained IQ. Sporadic performance might also be characteristic of patients with brain damage with diffuse cortical (Mittenberg, Hammeke, & Rao, 1989) or subcortical involvement (Godber, Anderson, & Bell, 2000). An analysis of the intrasubtest scatter can thus provide a type of information different from that obtained by merely looking at the quantitative-scaled scores. It should be noted, however, that research on this is equivocal given that

J. Ryan, Paul, and Arb (1999) were unable to find high subtest scatter on the Information subtest among patients who had documented retrieval difficulties.

Level V. Qualitative Analysis

The final step is to look at the content of responses, especially on Information, Vocabulary, Comprehension, and Similarities. Frequently, the presence of unique, highly personal, or unusual responses can suggest some important dimensions of an individual's intellectual or personality functioning (see Groth-Marnat et al., 2000; E. Kaplan et al., 1991, 1999). For example, some responses may reflect aggressive tendencies, concrete thinking, or unusual associations. A highly aggressive person might provide unusual responses on some of the Vocabulary items, or a person with paranoid personality characteristics might provide rigid, cautious, and legalistic responses. Similarly, impulsivity might be suggested by persons who quickly place incorrect blocks together on Block Design and then do not reflect on whether their designs were correct.

WECHSLER SUBTESTS

To interpret the Wechsler scales adequately, it is essential to understand the various abilities that each subtest measures. This section presents the different abilities involved in each of the 14 WAIS-III and 13 WISC-III subtests, followed by a discussion of their relevant features, including the possible meanings associated with high or low scores. Descriptions of the subtest abilities and data on factor loadings presented for most of the WISC-III subtests are derived from A. Kaufman (1994) and A. Kaufman and Lichtenberger (2000, 2002). Subtest abilities and factor loadings for the WAIS-III are based on research reviewed by A. Kaufman and Lichtenberger (1999, 2002) and Sattler (2001). Some citing of relevant and usually recent sources is also provided.

In keeping with the overall approach of this book, any interpretations suggested in the discussion of the subtests should be considered tentative. They are merely beginning possibilities that must be explored further and placed in a proper context. In addition, no subtest is a pure measurement of any single intellectual ability; rather, each represents a combination of skills. It is important to emphasize that a low or high score on a specific subtest can occur for a variety of reasons, which the examiner must consider in interpreting the overall profile. This section is most helpful only after practitioners are familiar with the subtest stimuli and administration procedure outlined in the WAIS-III and WISC-III manuals.

Verbal Scales

The Wechsler Verbal Scales assess an individual's proficiency in the following areas:

- The ability to work with abstract symbols.
- The amount and degree of benefit a person has received from his or her educational background.
- Verbal memory abilities.
- Verbal fluency.

The WAIS-III and WISC-III Verbal Scales are generally more subject to cultural influences, whereas the Performance Scales are considered to be somewhat more culture free. If an individual does significantly better (9 points or more for the WAIS-III or 12 points or more for the WISC-III) on the Verbal Scales compared with the Performance subtests, this difference may indicate a number of interpretative possibilities, including a relatively high level of education; a tendency toward overachieving; psychomotor slowing because of depression; difficulty working with practical tasks; deficits in performance abilities; poor visual-motor integration; a slow, deliberate, reflective work style that results in relatively lower scores on timed tests (but higher scores on verbal tests); or a quick, impulsive work style resulting in relatively more errors on Performance subtests (A. Kaufman, 1994; A. Kaufman & Lichtenberger, 1999, 2000, 2002; Sattler, 2001). In addition, persons from professional occupations, high educational attainment, and high IQs in general are likely to have quite high Verbal IQs. Also, psychiatric populations (5–6 point V > P discrepancy), persons with Alzheimer's disease, and persons with motor coordination problems tend to have higher verbal scores relative to their performance scores.

Studies with the WAIS-R have typically found that persons with unilateral right hemisphere lesions have, on average, a 9-point higher Verbal than Performance IQ (A. Kaufman, 1994; A. Kaufman & Lichtenberger, 1999, 2001, 2002; Reitan & Wolfson, 1993; Sattler, 2001). It is likely that future research will also find similar patterns with the WAIS-III. However, a V > P (e.g., depressed Performance IQ) should never be *diagnostic of* unilateral right hemisphere brain damage but rather *consistent with* this condition in some cases. It should be stressed that there is a complex interaction with a wide number of variables. A V > P effect is likely to be most pronounced among adult, educated (12+ years), Caucasian males with acute lesions who have strokes, tumors, or other focal lesions toward the posterior (versus anterior/frontal) regions. These variables have been extensively reviewed by Kaufman (1994) and Kaufman and Lichtenberger (2002) and are summarized in the following list:

- *Age.* Whereas the V > P effect has been clearly and consistently found for most adult populations, studies with children have been met with numerous contradictions. This is because there are a greater number of intervening variables for children, and their brains are more symmetrical and characterized by greater plasticity. Thus, neurological inferences related to Verbal-Performance discrepancies should *not* be made for children.
- *Education.* Because persons with higher education (and generally persons with higher IQs) typically score higher on Verbal subtests, a further lowering in performance abilities because of a right hemisphere lesion will serve to exaggerate the V > P discrepancy to an even greater extent. Persons from lower educational backgrounds often have higher Performance IQs relative to their Verbal IQs so that a lowering in their Performance IQ because of a right hemisphere lesion may either not produce the expected V > P effect, or the difference may not be as wide as for persons with higher educational attainment.
- *Race.* European American and African Americans are more likely to have the V > P discrepancy following right hemisphere damage than either Hispanics or

Native Americans. This is because Hispanics and Native Americans are more likely to have higher Performance than Verbal IQs before their injury or illness.

- *Gender.* The WAIS-R V > P discrepancy following right hemisphere lesions is more pronounced in males (13 points) than in females (7 points; A. Kaufman & Lichtenberger, 2002). This results partially from greater cerebral asymmetry in males. It is also possibly because of more verbally mediated strategies for Performance subtests by females, which serves to partially compensate for organically lowered performance abilities.
- *Recency of Lesion.* Acute (less than 12 months) unilateral right hemisphere lesions produce greater V > P effects than chronic lesions. This happens because, over time, patients are able to improve their performance abilities through both natural recovery of function and compensatory techniques. Even with chronic lesions, there is still an expected V > P discrepancy, but it is not as extreme as for acute lesions.
- *Type and Location of Lesion.* Especially right hemisphere strokes, but also tumors and, to a lesser extent, right temporal lobe epilepsy result in the expected V > P effect. Frontal lobe lesions have little effect on V – P differences, whereas posterior lesions do result in the expected V > P discrepancy.

Vocabulary

The Vocabulary subtest includes the following abilities or traits:

- Language development.*
- Word knowledge.*
- General verbal intelligence.
- Language usage and accumulated verbal learning ability.
- Rough measure of the subject's optimal intellectual efficiency.
- Educational background.
- Range of ideas, experiences, or interests that a subject has acquired.

The Vocabulary subtest is a test of accumulated verbal learning and represents an individual's ability to express a wide range of ideas with ease and flexibility. It may also involve the person's richness of ideas, long-term memory, concept formation, and language development. Vocabulary is noteworthy in that it is the most reliable Verbal subtest (WAIS-III test-retest reliability = .91; WISC-III test-retest reliability = .89) and, like Information, it is highly resistant to neurological deficit and psychological disturbance (Lezak, 1995; Reitan & Wolfson, 1993). Although the Vocabulary subtest holds up with age, it tends to fall off with those people for whom visual-spatial skills are far more important than verbal abilities. Vocabulary generally reflects the nature and level of sophistication of the person's schooling and cultural learning. Vocabulary is primarily dependent on the wealth of early educational environment, but it is susceptible to

* Abilities followed by an asterisk indicate specific abilities and traits strongly associated with the subtest under discussion.

improvement by later experience or schooling. It is the least variable of all the subtests, and subtest scores below the Vocabulary level sometimes imply a drop of efficiency in that function. Vocabulary is the best single indicator of general intelligence, with 69% of its variance accounted for by *g* on the WAIS-III and 80% of its variance accounted for by *g* on the WISC-III. Because of its high degree of stability, Vocabulary is often used as an indicator of a person's intellectual potential and to make an estimate of their premorbid level of functioning (see more precise methods in "Assessing Brain Damage" section).

The Vocabulary responses are similar to Comprehension and Similarities in that a qualitative analysis often provides useful information relating to the examinee's thought processes, background, life experiences, and response to frustration. It is often important to explore incorrect responses to determine whether they were guesses, clang associations (e.g., "ponder" meaning "to pound" or "assemble" meaning to "resemble"), concrete thinking, bizarre associations, or overinclusive reasoning. Even when a response is correct, a consideration of the style used to approach the word and specific content can be helpful.

High scores suggest high general intelligence and indicate that the examinee can adequately recall past ideas and form concepts relating to these ideas. Persons with high scores have a wide range of interests, a good fund of general information, and may have high needs for achievement. Clinical populations who score high on Vocabulary may use compulsive or intellectualizing defense mechanisms. Low scores suggest a limited educational background, low general intelligence, poor language development, lack of familiarity with English, and/or poor motivation.

Similarities

- Logical abstract reasoning.*
- Verbal concept formation or conceptual thinking.
- Distinguishing essential from nonessential details.
- Associative ability combined with language facility.

The Similarities subtest requires verbal concept formation and abstract reasoning ability. These functions mediate for the individual an awareness of the belonging-togetherness of objects and events of the day-to-day world. An essential aspect of adjusting to one's environment is the use of these abilities to clarify, reduce, and classify the style and manner to which a response is made. Inductive reasoning is required as the examinee must move from particular facts to a general rule or principle. Implicit in the test is the ability of individuals to use long-term memory and to apply elegant expressions in their responses. The more precise and abstract the expression, the higher the score, which indicates that verbal fluency is an important determinant. Correct responses to the last few items indicate a particularly high level of abstraction. Individuals with a good ability for insight and introspection tend to perform highly on this subtest; thus, it may be used as an indicator of favorable prognosis for psychotherapy. Scores decrease significantly in schizophrenics, rigid or inflexible thinkers, and patients with senile conditions. Examiners can, therefore, use this subtest to gain further information regarding the nature of an examinee's idiosyncratic or pathological form of concept formation.

High scorers show good verbal concept formation, which, if unusually high, may reflect intellectualizing tendencies. Low scorers show poor abstraction abilities, literalness, and inflexible thinking. The Similarities subtest in adult protocols is the most sensitive subtest to left hemisphere lesions, particularly lesions to the left temporal and/or left frontal regions (Dobbins & Russell, 1990).

Arithmetic

- Computational skill.*
- Auditory memory.
- Sequencing ability.
- Numerical reasoning and speed of numerical manipulation.
- Concentration and attention/low distractibility.
- Reality contact and mental alertness; that is, active relationship to the outside world.
- School learning (earlier items)/acquired knowledge.
- Logical reasoning, abstraction, and analysis of numerical problems (later items).

The Arithmetic subtest requires a focused concentration as well as basic mathematical skills and an ability to apply these skills. The skills required to complete this test are usually acquired by the time a person reaches junior high school; therefore, low scores are more likely to be the result of poor concentration. Arithmetic is likely to be more challenging and stressful than tests such as Information and Vocabulary, both because the task itself is more demanding and because the test is timed. Thus, persons who are susceptible to the disruptive effects of anxiety are likely to be adversely affected. However, examiners may want to establish whether the person simply lacked the necessary skills or had difficulty concentrating. This can be assessed by giving the person previously missed items a second time but allowing the use of paper and pencil without a time limit. Under these circumstances, persons with adequate mathematical knowledge who are distractible should be able to complete the items correctly.

Individuals from higher socioeconomic backgrounds, obedient teacher-oriented students, and persons with intellectualizing tendencies usually do well on this subtest. A helpful formula is that Information plus Arithmetic equals school achievement. Because numbers come from the outside environment and create rule and direction, some individuals react rebelliously. This is particularly true for antisocial personalities. Histrionic personalities, who do not readily accept outside direction and generally refuse to take responsibility for their behaviors, may likewise do poorly. This is not to suggest that lowered Arithmetic scores are diagnostic of these clinical groups, but rather, that this lowering may at times be consistent with the way these individuals interact with their environment.

High scorers show alertness, capacity for concentration, freedom from distractibility, and good short-term auditory memory and may use intellectualizing defenses. Low scorers show poor mathematical reasoning, lack of capacity to concentrate, distractibility, and poor auditory short-term memory. A poor educational background in which adequate mathematical skills have not been developed can also account for lowered performance.

Digit Span

- Immediate rote recall.*
- Reversibility; ability to shift thought patterns (from digits forward to digits backward).*
- Concentration and attention.
- Auditory sequencing.
- Rote learning.

Digit Span is considered to be a test of short-term memory and attention. The subject must recall and repeat auditory information in the proper sequence. Bannatyne (1974) has further described this as "auditory vocal sequencing memory." Correct responses require a two-step process. First, the information must be accurately received, which requires attention and encoding. Persons who are easily distractible have difficulty in this phase. Second, the examinee must accurately recall, sequence, and vocalize the information. Persons who can perhaps receive the information correctly may still have difficulty at this phase if they have short-term memory difficulties because they cannot hold the memory trace long enough. Sometimes, the previous digit is forgotten as they are attempting to vocalize a present one. Whereas Digits Forward is a simpler, more straightforward task requiring rote memory, Digits Backward is more complex. The examinee must usually hold the memory longer and also transform it before making a restatement. Thus, a good performance on Digits Backward is likely to reflect a person who is flexible, can concentrate, and is tolerant of stress. High Digits Backward scores may also involve the ability to form, maintain, and scan visual mental images formed from the auditory stimulus (Lezak, 1995; Wielkiewicz, 1990).

Passive, anxiety-free individuals seem to do best on this test. It requires an effortless and relatively unhampered contact with reality, which is characterized by open receptivity to incoming information. Performance is greatly hampered by increased anxiety or tension, and the Digit Span subtest is considered the most susceptible to the effects of anxiety. In addition to Digit Span, the other subtests that are sensitive to the effects of anxiety are Arithmetic, Digit Symbol-Coding, and Letter-Number Sequencing (WAIS-III). Collectively, these three subtests form the WAIS-III Working Memory Index and are (along with the Processing Speed subtests) sensitive tests to brain damage, mental retardation, and learning disabilities (Lezak, 1995; Psychological Corporation, 1997). Similarly, the Digit Span subtest (and Arithmetic) is included in the WISC-III Freedom from Distractibility Index, which is also sensitive to the effects of learning disabilities, ADHD, brain damage, and mental retardation (Bannatyne, 1974; A. Kaufman, 1994).

Persons who score high have good auditory short-term memory and excellent attention and may be relatively unaffected by stress and anxiety. However, just because a person has good short-term auditory memory for digits does not necessarily mean that his or her memory for more complicated information, such as music or verbally relevant information, is also good. These more complex features of memory may have to be assessed by other means. When Digits Backward is longer than Digits Forward, this rare event (3% to 10% of children's protocols; Wechsler, 1991; .9% of adult profiles, Psychological Corporation, 1997) suggests that the individual has excellent numerical

abilities. Low scores on Digit Span indicate difficulty concentrating, which may be the result of anxiety or unusual thought processes. A large discrepancy (5 digits) in favor of Digits Forward versus Digits Backward can suggest the presence of an organic deficit, particularly if the overall backward Digit Span score is below scores for tests such as Information and Vocabulary. Whereas Digits Forward is fairly stable and resistant to deterioration, Digits Backward is a far more difficult task and is quite sensitive to deterioration (see subsection on estimating premorbid IQ in the "Assessing Brain Damage" section). Whereas Digits Forward is more likely to be lowered by left hemisphere lesions, lowered Digits Backward is more consistent with either diffuse or right frontal involvement (Lezak, 1995; Swierchinsky, 1978). Lowered performance for both Digit Span backward and Digit Symbol occur with the diffuse damage associated with exposure to solvents (Groth-Marnat, 1993; Morrow, Furman, Ryan, & Hodgson, 1988).

Information

- Range of general factual knowledge.*
- Old learning or schooling.
- Intellectual curiosity or urge to collect knowledge.
- Alertness to day-to-day world.
- Long-term memory.

The Information subtest samples the type of knowledge that average persons with average opportunities should be able to acquire. This knowledge is usually based on habitual, overlearned material, particularly in the case of older children and adults. Both Information and Vocabulary are highly resistant to neurological deficit and psychological disturbance (Lezak, 1995; Reitan & Wolfson, 1993) and are two of the most stable subtests. Because of this stability, Wechsler referred to them as "hold" tests as opposed to "no-hold" tests, which he theorized are more sensitive to deterioration and such situational variables as anxiety and fatigue (i.e., Arithmetic, Digit Symbol-Coding, Block Design). Furthermore, both these subtests are good measures of general intelligence and are highly correlated with educational level (A. Kaufman et al., 1988) and WAIS-III and WISC-III Full Scale IQs. Research has shown that the earlier WAIS-R Information and Vocabulary subtests have predicted college grade point average as accurately as well-established college aptitude tests (Feingold, 1983). It is for these reasons that Information (along with Vocabulary and Arithmetic) is included in Bannatyne's Acquired Knowledge category. It also loads most strongly (.84) on the Verbal Comprehension factor.

Although performance on the Information subtest involves remote memory and alertness to the environment, it is influenced only to a small extent by conscious effort and is believed to be only minimally affected by factors such as anxiety. To score well, the individual must have been exposed to a highly varied past environment, have an intact long-term memory, and possess a wide range of interests.

A high score on this subtest suggests that the examinee has good long-term memory, cultural interests, strong educational background, positive attitude toward school, good verbal ability, and possibly intellectualization as his or her most frequently used defense mechanism. Low scorers may show superficiality of interests, lack of intellectual curiosity, cultural deprivation, or lack of familiarity with Western (primarily American)

culture (however, note the availability of numerous foreign country adaptations). Failing initial easy items combined with success on more difficult ones (high intrasubtest variability; see Level IV procedure) may suggest difficulties with retrieval, although research substantiating this hypothesis has been equivocal (E. Kaplan et al., 1991; Mittenberg et al., 1989; J. Ryan & Paul, 1999). High intrasubtest scatter may also suggest the possibility of malingering or poor motivation.

Comprehension

- Demonstration of practical knowledge.*
- Social maturity.*
- Knowledge of conventional standards of behavior.*
- Ability to evaluate past experience; that is, proper selection, organization, and emphasis of facts and relationships.*
- Abstract thinking and generalization (later items only).*
- Social judgment, common sense, or judgment in practical social situations.
- Grasp of social milieu; for example, information and knowledge of moral codes, social rules, and regulations.
- Reality awareness, understanding, and alertness to the day-to-day world.

Comprehension has often been considered to reflect the extent to which an examinee adheres to conventional standards, has benefited from past cultural opportunities, and has a well-developed conscience. However, formal studies have generally not supported a relationship between Comprehension and various measures of social intelligence (see Beebe, Pfiffner, & McBurnett, 2000). Comprehension is also, at least in part, a test of information, which is supported by its high correlation (low- to mid-70s, depending on age) with the Information and Vocabulary subtests. Comprehension involves an adaptive response by the individual to a situation that requires him or her to select the most efficient way of dealing with a specific problem. The examinee not only must possess relevant information but also must appropriately use this information for decision making. In this sense, the Comprehension subtest goes one step beyond the degree of complexity and synthesis required for the Information subtest. Like Vocabulary and Information, it measures general verbal ability—66% of its WAIS-III variance and 42% of its WISC-III variance are attributed to the Verbal Comprehension factor. The examinee must not only have the necessary information, but also apply it in a coherent, problem-oriented manner. Thus, a Comprehension score significantly below the Information score suggests that an examinee is not effectively using his or her knowledge.

In assessing an examinee's responses, it can be important to distinguish between actually dealing with the material to develop an original response and merely repeating overlearned concepts. For example, parroting answers to "forest," "parole system," or the proverbs does not indicate full comprehension and may simply be based on past experience rather than on accurate problem solving, good judgment, or abstract reasoning. Thus, basic rule-of-thumb answers can significantly increase the total number of correct responses. However, in the later items, a correct response requires higher-level problem solving, and these items, therefore, can still be a good measure of general intelligence instead of merely rote memorization.

Personality variables, especially those relating to judgment, are important areas to consider in this subtest. In particular, poor levels of adjustment can lower scores on Comprehension. Clinicians should note the pattern of responses, clichés, literalness, and any circumscribed responses. In contrast, good judgment involves the ability to engage in discriminative activity. Failure on the easy items indicates impaired judgment, even though later, more difficult items are passed. It is important to note emotional implications on this subtest because emotional responsiveness influences the way a person evaluates environmental events. For example, individuals who are highly analytical and use these analytical abilities to avoid emotions may have difficulty understanding the social components of situations as presented in Comprehension.

High scorers show reality awareness, capacity for social compliance, good judgment, and emotionally relevant use of information. Low scorers, especially if they have four or more subscale points below Vocabulary, might have poor judgment, impulsiveness, and hostility against their environment. Mentally disturbed persons often do poorly on Comprehension, which may be the result of disturbed perceptions, idiosyncratic thinking, impulsiveness, or antisocial tendencies.

Letter-Number Sequencing (WAIS-III only)

- Auditory short-term memory.
- Sequencing ability.
- Concentration and attention.

A good performance on Letter-Number Sequencing suggests that the person has good sequencing, attention, and concentration. It requires him or her to attend to a series of letters and numbers that have been read to him or her, hold them in memory, manipulate them into a new order, and repeat the new sequence. When combined with Arithmetic and Digit Span, it forms the Working Memory Index, but it is not used to calculate any of the IQs. Letter-Number Sequencing (along with Digit Span) is a subtest, which is also included on the Wechlser Memory Scale-III.

Psychometrically, Letter-Number Sequencing is good to adequate. Test-retest reliability has been found to range between .70 to .80, the SEM is 1.30, and it has a factor loading of .62 with the Working Memory Index.

Performance Scales

The Performance scales reflect:

- The individual's degree and quality of nonverbal contact with the environment.
- The ability to integrate perceptual stimuli with relevant motor responses.
- The capacity to work in concrete situations.
- The ability to work quickly.
- The ability to evaluate visuo-spatial information.

The Performance subtests are generally less affected by educational background than are the Verbal scales. If an individual does significantly (.05 level) better (9 points or

more on the WAIS-III, 12 or more points on the WISC-III) on the Performance scales than on the Verbal subtests (P > V), this may indicate a number of interpretive possibilities, including superior perceptual organizational abilities, ability to work under time pressure, a tendency toward low academic achievement, possible acting out (juvenile delinquency), an individual who could be described as a doer rather than a thinker, a person from a relatively low socioeconomic background, presence of a language deficit, poorly developed auditory conceptual/processing skills, or that immediate problem solving is better developed than problem solving based on accumulated knowledge.

A number of studies, primarily with the WAIS/WAIS-R have found that a higher Performance than Verbal IQ (P > V) is consistent with unilateral left hemisphere lesions (A. Kaufman & Lichtenberger, 2002). There is, however, a complex relation between a number of relevant variables and P > V for unilateral left lesion patients. One issue is that the average Verbal IQ superiority of 4 points for unilateral left lesion patients across studies is not nearly as pronounced as the 9-point average for V > P with right hemisphere lesions.

Because the P > V effect is not as strong as the V > P discrepancy found with unilateral right hemisphere lesions, interpretations need to be quite tentative. In general, P > V discrepancies are most likely to occur for adult male patients with low educational attainment who have lesions in the posterior (versus frontal) regions (see A. Kaufman & Lichtenberger, 2002). These variables can be summarized as follows:

- *Age.* The P > V difference for left lesion adults is relatively small but has been found not to occur for children. Therefore, inferences regarding lateralization should be restricted to adults and adolescents.
- *Gender.* The laterality effect for P > V following unilateral left hemisphere lesions has been found to be greater for males (6 points) than for females (only 1 point; A. Kaufman & Lichtenberger, 2002; see also previous section on gender for V > P following right hemisphere lesions).
- *Education.* Individuals having less than a high school education generally score 2 to 3 points higher on their Performance IQ than Verbal IQ. Clinically, this means that persons with low educational attainment are more likely to have even greater P > V following unilateral left hemisphere lesions than persons with more education.
- *Type and Location of Lesion.* Posterior left lesions are likely to show the expected P > V difference. Frontal lesions, no matter what the cause, are not likely to demonstrate any V > P differences. Left hemisphere strokes tend to produce the clearest P > V effect and, to a lesser extent, left temporal lobe epilepsy. Left hemisphere tumors, as well as the relative recency of the lesion (acute versus chronic), have little effect on V > P discrepancies.

A further consideration related to P > V difference is that research with the WAIS-R/WISC-III indicates that certain population groups are likely to score higher on Performance subtests. In particular, children, adolescents, and adult Native Americans and Hispanics (especially if bilingual) have Performance scores that can be an average of nearly 15 points above their Verbal scores. As a result, Wechsler Intelligence scale interpretation, especially if related to Verbal or Full Scale IQs, should be made with extreme

caution (if at all). Instead, the Verbal and Performance IQs should be considered separately. Additional correlates of P > V are autism, mental retardation, learning disabilities, illiteracy, delinquency, conduct disorder or psychopathy, bilingual populations, and individuals from occupations (especially blue-collar) emphasizing visual-spatial skills. Possible explanations for these differences include the challenges involved in learning two languages, the level to which the test instructions have been understood, attitudes and experience working within time constraints, degree of cerebral lateralization, and cultural or subcultural differences (i.e., extent that nonverbal communication is emphasized). Each of these correlates should be taken into account when making interpretations related to lateralization of brain lesions or any of the other possible interpretations consistent with P > V discrepancies.

Picture Completion

- Visual alertness.*
- Visual recognition and identification (long-term visual memory).*
- Awareness of environmental detail; reality contact.
- Perception of the whole in relation to its parts; visual conceptual ability.
- Ability to differentiate essential details from nonessential details.
- Visual concentration combined with an ability to visually organize material.

The Picture Completion subtest is a measure of visual concentration and is a nonverbal test of general information. It involves discovering consistency and inconsistency by paying close attention to the environment and accessing remote memory. It is dependent on, and also draws on, an individual's experience with his or her culture. Thus, a person who is unfamiliar with common features of American/Western society may make errors because of a lack of experience rather than a lack of intelligence. A person will also make errors if he or she is unable to detach himself or herself emotionally from the material, thereby making accurate discriminations difficult. For example, passive, dependent personalities might make errors because they notice the absence of people controlling the actions in the pictures. Typical responses might be that "there's nobody holding the pitcher," "there are no people rowing the boat," or "there's no flagpole." Sometimes negative, inflexible, oppositional individuals state that there is nothing missing in the pictures.

High scorers are able to recognize essential visual information, are alert, and demonstrate good visual acuity. Low scores indicate poor concentration and inadequate visual organization. Impulsiveness can often produce lowered performance because the examinee may make a quick response without carefully analyzing the whole picture.

Digit Symbol-Coding/Coding

- Psychomotor speed.*
- Ability to follow directions.*
- Clerical speed and accuracy.*
- Visual short-term memory.*
- Ability to follow directions.*

- Paper-pencil skills.*
- Ability to learn an unfamiliar task; capacity for learning and responding to new visual material.
- Some degree of flexibility; ability to shift mental set.
- Capacity for sustained effort, attention, concentration, and mental efficiency.
- Associative learning and ability to imitate newly learned visual material.
- Sequencing ability.

Visual-motor integration is implied by good performance on Digit Symbol-Coding. However, the most important functions necessary for a high score are psychomotor speed combined with good recall for the symbol-digit pairs. This test involves appropriately combining the newly learned memory of the digit with the symbol, as well as adequate spatial-motor orientation, followed by executing the half-habituated activity of drawing the symbol. The subtest also requires the ability to learn an unfamiliar task, accuracy of eye-hand coordination, attentional skills, short-term memory, and the ability to work under pressure. This is a delicate and complex interaction, which can be disturbed because of difficulties with any of the preceding skills. In contrast to Vocabulary, which is a highly stable subtest, Digit Symbol is extremely sensitive to the effects of either organic or functional impairment. In particular, depressed patients and patients with brain damage have a difficult time with this subtest. It is also the subtest that is most influenced by age. For example, a raw score required to achieve a subscale score of 10 for the 70- to 74-year-old group would obtain a subscale score of only 6 when compared with the 20- to 34-year-old reference group.

Digit Symbol-Coding pairs with Symbol Search to form the Processing Speed Index. Digit Symbol-Coding is a fair measure of *g* for the WAIS-III (35% of its variance) but only a poor measure of *g* for the WISC-III (20% of its variance). It has ample subtest specificity for both the WAIS-III and WISC-III.

Because visual-motor coordination (particularly visual acuity and motor activity) is implied, it is not surprising to find that those individuals with high reading and writing experience are among the high scorers. Functions that are implicit in the task are rapid visual, spatial, and motor coordination, as well as the executive action of drawing the symbol. Because this task requires sustained attention and quick decision making, anxious hesitancy, obsessiveness, deliberation, and perfectionism significantly lower scores. This difficulty might be somewhat counteracted by informing persons who appear perfectionistic and reflective that they need only make their responses legibly but not perfectly. Persons who are extremely competitive but also become highly anxious in competitive situations may also be adversely affected. Not only can Digit Symbol-Coding scores be lowered by anxiety, but also the psychomotor slowing found in depressive states or the confused orientation of schizophrenics likewise produces a decrease in performance. Thus, a rough index of the severity of a person's depression can be assessed by comparing the relative lowering of Digit Symbol-Coding with other more stable subtests. Of particular significance is that Digit Symbol-Coding is one of the most sensitive subtests to the effects of any type of organic impairment (Lezak, 1995; Psychological Corporation, 1997; Reitan & Wolfson, 1993), and it tends to be one of the lower scores found in learning-disabled individuals (Bannatyne, 1974; Groth-Marnat,

2002; A. Kaufman, 1994). Even with minimal brain damage, Digit Symbol-Coding is still likely to be the lowest subtest overall (Lezak, 1995; Reitan & Wolfson, 1993). In addition, patients with rapidly growing tumors are more likely to have lower scores than those with slow-growing tumors (Reitan & Wolfson, 1993).

Because Digit Symbol-Coding requires such a diverse range of abilities, high or low scores can potentially indicate a wide number of possibilities. This means that clinicians need to work particularly hard to extract the significance of scores by integrating scores with other relevant measures, behavioral observations, and medical/personal history. The WAIS-III has included two optional procedures to help parcel out whether an examinee's score was attributable primarily to visual memory, graphomotor speed, or a combination of both. The first procedure, Incidental Learning, assesses how intact visual memory is by first requesting patients to recall as many of the digit-symbol pairs as possible and, second, to simply recall as many symbols as possible (without the associated numbers). These two related tasks are untimed. In contrast, Digit Symbol-Copy assesses graphomotor speed by presenting the examinee with a series of symbols and then requests that he or she write down as many of the symbols as possible in boxes directly under the symbol. The examinee is given 90 seconds to write down as many of the symbols as possible. Various combinations of high and low scores can help to understand the underlying processes involved with Digit Symbol-Coding. For example, if a client did poorly on Digit Symbol-Coding and Incidental Learning was high (e.g., good visual memory) but Digit Symbol-Copy was low (e.g., slowed graphomotor speed), it suggests the reason for the poor performance was slow graphomotor speed.

High scorers potentially have excellent visual-motor ability, mental efficiency, capacity for rote learning of new material, and quick psychomotor reactions. Lower scorers may have reduced capacity for visual associative learning, impaired visual-motor functioning, and poor mental alertness.

Block Design

- Analysis of whole into component parts.*
- Spatial visualization.*
- Nonverbal concept formation.
- Visual-motor coordination and perceptual organization.
- Capacity for sustained effort; concentration.
- Visual-motor-spatial coordination; manipulative and perceptual speed.

The Block Design subtest involves nonverbal problem-solving skills because it emphasizes analyzing a problem into its component parts and then reintegrating these parts into a cohesive whole. The examinee must apply logic and reasoning in a manner that will solve spatial relationship problems. As a test of nonverbal concept formation, Block Design demands skills in perceptual organization, spatial visualization, and abstract conceptualization. The Block Design subtest is sturdy and reliable, correlating highly with general intelligence, and is not likely to be lowered except by the effects of depression or organic impairment. Also it has been found to relate to everyday

measures of spatial abilities (Groth-Marnat & Teal, 2000). To perform well, examinees must be able to demonstrate a degree of abstraction that is free from literal concreteness. They must also make a distinction between part and whole by demonstrating both analytic and synthetic skills. This test involves an ability to shift the frame of reference while maintaining a high degree of flexibility. The examinee must also be able to inhibit his or her impulsive tendencies and to persist in a designated task.

An important feature of Block Design is that it enables an examiner to actually observe the examinee's response. Some subjects are easily discouraged and give up, while others insist on completing the task even if they have to work beyond the time limit. In approaching the task, one subject might impulsively place the blocks together in a nonrandom sequence, whereas another subject might demonstrate a meticulous sequential style, thereby revealing preferences for either a holistic simultaneous or a more sequential problem-solving style. Additional observations can reveal factors such as hand preference, motor coordination, speed of information processing, frustration tolerance, and ability to benefit from feedback. A highly reflective or compulsive style can lower scores because of the resulting extended time for completing the task. Placing blocks outside the 2×2 or 3×3 configuration is a further behavioral observation that reflects poor visuos-patial skills (J. H. Kramer, Kaplan, & Huckeba, 1999). Thus, potentially valuable information can be obtained by observing and recording differences in solving the Block Design tasks.

Block Design is also a nonverbal, relatively culture-free test of intelligence. It is reliable in that it correlates highly with general intelligence (approximately 52% of its variance may be attributed to *g*), but it has a relatively low correlation with education. Thus, the Block Design subtest is only minimally biased by an examinee's cultural or educational background. Block Design scores can, therefore, be an important tool in assessing the intellectual potential of persons from divergent cultural and intellectual backgrounds.

Block Design is an excellent indicator of right hemisphere brain damage and is especially sensitive to right parietal lesions (Lezak, 1995; Reitan & Wolfson, 1992, 1993). Right lesion patients tend to make errors because they might distort the designs, misperceive aspects of them, or become disoriented when attempting to complete them. In contrast, left lesion patients, particularly if the lesion is in the parietal lobe, are not nearly as likely to have a poor Block Design score. However, when they do, it is likely to be expressed in design simplification, confusion, and a concrete approach to reproducing the design (Lezak, 1995). Inattention (neglect) can be reflected by the examinee's failing to complete the right or left portion of the design. For example, only six or seven of the blocks might be used when attempting to complete a nine-block design (Lezak, 1995). Block Design is typically one of the lowest subtest in Alzheimer's patients. It is sensitive to the early phases of the disease and thus can be useful in differentiating between Alzheimer's and pseudodementing conditions such as depression (Fuld, 1984; La Rue & Jarvik, 1987).

High scorers show a good capacity for visual-spatial perception, visual-motor speed, a good ability to concentrate, and excellent nonverbal concept formation. Low scores suggest poor perceptual abilities, difficulties with visual integration, and problems in maintaining a sustained effort.

Matrix Reasoning (WAIS-III only)

- Visual-spatial reasoning.
- Abstract reasoning.
- Visual organization.
- Simultaneous processing of visual-spatial information.
- Analysis of wholes into component parts.

High scores on Matrix Reasoning suggest good visual information processing and nonverbal abstract reasoning skills. It is combined with Picture Completion and Block Design to form the Perceptual Organization Index. Matrix Reasoning is untimed and is, therefore, useful for persons from older age groups who might do poorly on some of the other timed tests. It also does not penalize those who have a reflective, cautious problem-solving style. Matrix Reasoning is relatively culture free and requires only a minimal amount of visual motor-coordination because the subject merely points to the correct response. Conceptually, Matrix Reasoning is similar to the Halstead Reitan Category Test and Raven's Progressive Matrices. However, future studies will need to determine the nature and degree of correspondence between these measures.

One of the rationales for Matrix Reasoning was to develop a visual-spatial subtest with good psychometric properties that could replace the psychometrically poor Object Assembly subtest. In many ways, this has been realized as Matrix Reasoning has been found to have test-retest stabilities ranging from .75 to .81, SEM of .97, a correlation with the Full Scale IQ of .75, and a factor loading of .61 on the Perceptual Organization Index. It is one of the best performance subtest measures of *g* (52% of its variance can be attributed to *g*). In contrast, Object Assembly has poorer psychometric properties with particular concerns related to its lack of stability (SEM = 1.66). As a result, Object Assembly is now an optional WAIS-III subtest.

High scores might indicate good nonverbal abstract reasoning abilities, a preference for simultaneous processing of information, and excellent visual information processing. Low scores might suggest low visual concept formation, poor or, at least, rigid visual reasoning, or poor concentration. Negativism might be indicated if the examinee seems unmotivated and replies with wording such as "none of them match."

Picture Arrangement

- Planning ability (comprehending and sizing up a total situation).*
- Anticipation of consequences.*
- Temporal sequencing and time concepts.*
- Accurately understanding nonverbal interpersonal situations.
- Ability to comprehend a total situation and evaluate its implications.
- Visual organization and perception of essential visual cues.
- Speed of associating and planning information.

The Picture Arrangement subtest is primarily a test of the ability to plan, interpret, and accurately anticipate social events in a given cultural context. Thus, an individual's cultural background can affect his or her performance on the test; normal subjects

with poor or different cultural backgrounds often do poorly. This means that scores derived from such persons should be treated with caution. Wechsler (1958) stated that the test requires an examinee to use general intelligence in nonverbal social situations. In fact, each of the items requires a person to respond to some practical interpersonal interaction. Solving the correct sequence also requires at least some sense of humor. However, interpretive caution should be exercised because most research has not supported relationships between Picture Arrangement and measures of social intelligence (Beebe et al., 2000; Lipsitz et al., 1993). Both Picture Arrangement and Block Design are measures of nonverbal intelligence. However, Picture Arrangement is far more dependent on cultural variables than is Block Design. Picture Arrangement also requires the person to grasp or "size up" the complete situation before proceeding to a correct response. In contrast, persons can achieve good scores on Block Design by approaching the task in small segments and then contrasting their performance on each segment with the whole design.

Picture Arrangement is somewhat sensitive to the effects of brain damage, especially for those injuries that disrupt nonverbal social skills (Golden, 1979; Lezak, 1995). An unusually low Picture Arrangement score in a protocol in which there is little difference between Verbal and Performance IQs implies an organic impairment consistent with a static lesion to the right anterior temporal lobe (Reitan, 1974a; Reitan & Wolfson, 1993). More generalized right hemisphere lesions are likely to lower not only scores on Picture Arrangement, but also performance on Block Design and Object Assembly (Russell, 1979). There is also some evidence that patients with frontal lobe impairment do poorly on Picture Arrangement because of their tendency to respond impulsively and without considering the entire problem (Walsh, 1994).

Two approaches can be followed to obtain additional qualitative information from Picture Arrangement. The first is to observe and record how the person attempts to solve the problem. Does the client carefully consider the overall problem or rather impulsively begin altering the cards? Is the client easily discouraged or does he or she demonstrate a high degree of persistence? After the entire subtest has been completed, an examiner may also want to obtain a subject's description of the stories related to the pictures. This might be initiated by simply asking the examinee to "Tell me what is happening in the pictures" or "Make up a story about the cards." The following questions are especially important: Are the stories logical, fanciful, or bizarre? Are they original or rather stereotyped and conventional? Do examinees reveal any emotional attitudes relating either to themselves or to their interpersonal relationships? Were errors the result of incorrectly perceiving specific details or rather of neglect in even considering certain details? Did the examinee consider all the different relationships in the pictures or were important aspects omitted?

The previous information on Picture Arrangement applies primarily to the WAIS-III rather than the WISC-III because a substantial amount of extra credit for speed was given for the WISC-III revision of Picture Arrangement. It relates quite closely to the Processing Speed factor (along with Coding and Symbol Search; Hishinuma & Yamakawa, 1993; Wechsler, 1991). The practical implication is that WISC-III interpretation of Picture Arrangement scores should emphasize the speed component above or, at least in the context of, Picture Arrangement's other aspects (e.g., understanding nonverbal interpersonal situations).

Persons who score high on Picture Arrangement are usually sophisticated, have a high level of social intelligence, and demonstrate an ability to quickly anticipate the consequences of initial acts. Low scorers may have a paucity of ideas, difficulty planning ahead, slow processing of information, a poor sense of humor, difficulty in interpersonal relationships, and poor rapport.

Symbol Search

- Speed of visual search.*
- Speed of processing information.
- Planning.
- Encoding information in preparation for further processing.
- Visual-motor coordination.
- Learning ability.

Symbol Search was designed to be as pure a test as possible of information-processing speed. It pairs nicely with Digit Symbol-Coding because, conceptually, they assess similar areas, as is more formally indicated by relatively high correlations (WAIS-III, .65; WISC-III, .53) between the two subtests. Together, they form the Processing Speed factor. Symbol Search is psychometrically a relatively good subtest. Test-retest over a 2- to 12-week interval was .79 for the WAIS-III and .76 for the WISC-III. It correlates relatively highly with both Full Scale (WAIS-III, .66; WISC-III, .56) and Performance (WAIS-III, .69; WISC-III, .58) IQs.

High scores suggest that the individual can rapidly absorb information as well as integrate and respond to this information. In addition, it suggests good levels of visual-motor coordination, short-term visual memory, planning, general learning, and a high level of attention and concentration. Low scores suggest slow mental processes; visual-perceptual difficulties; possibly poor motivation and/or anxiety; difficulties with short-term visual memory; and a reflective, perfectionistic, or obsessive problem-solving style.

Object Assembly

- Ability to benefit from sensory-motor feedback.*
- Anticipation of relationships among parts.*
- Visual-motor organization.
- Simultaneous (holistic) processing.
- Synthesis; putting things together in a familiar configuration.
- Ability to differentiate familiar configurations.
- Manipulative and perceptual speed in perceiving the manner in which unknown objects relate to each other.

Object Assembly is a test of motor coordination and control, as are Digit Symbol-Coding and Block Design. It measures the ability to differentiate familiar configurations, and it also involves some anticipation and planning. However, scores are subject to a high degree of fluctuation, primarily because of the potential for accidentally fitting together parts. A related area that may create some confusion is that persons who are in

the lower ranges of intelligence (60 to 75) sometimes do quite well, whereas persons with above-average IQs can do quite poorly. The preceding difficulties have resulted in only moderate test-retest reliabilities (WAIS-III, .76; WISC-III, .64 to .71). In addition, Object Assembly is only a moderate measure of general intelligence (WAIS-III, 38%, and WISC-III, 44% of its variance may be attributed to *g*) and is not highly correlated with Full Scale IQ scores (WAIS-III, .59; WISC-III, .58). Furthermore, its correlation with other subtests is generally low. This is why it became an optional subtest for the WAIS-III. Because it is psychometrically one of the poorest subtests, scores should be treated with caution. In addition, it generally lacks a sufficient amount of subtest specificity for adequate interpretation of the test's underlying abilities.

Despite these difficulties, an advantage of Object Assembly is that, as with Block Design and Picture Arrangement, an examiner can directly observe a person's problem-solving style and reactions to success or failure. The test presents an "open" situation, and those who can work freely in this context usually do well. However, those with rigid visual organizations stick with one clue without allowing themselves to change their frame of reference. This inflexibility is often seen with people who are obsessive-compulsive. On the other hand, a flexible visual organization permits a rapid integration of new clues and an adaptation of these clues toward completing the task. The same observations relevant for Block Design are appropriate for Object Assembly. These include persistence, concentration, hand preference, frustration tolerance, speed of processing information, reflectivity, impulsiveness, ability to benefit from feedback, and preference for a simultaneous versus a sequential problem-solving style. In particular, an overly cautious, reflective, and/or obsessive approach is likely to lower performances because of the loss of bonus points resulting from their slow completion of the task.

Persons scoring high on Object Assembly show good perceptual-motor coordination, have superior visual organization, and can maintain a flexible mental outlook. Low scorers show visual-motor disorganization, concreteness, and difficulties with visual concept formation. Like Block Design, Object Assembly is sensitive to right, especially right posterior, lesions (Lezak, 1995; Reitan & Wolfson, 1993). However, given the test's inadequate test specificity and low reliabilities, these interpretations should be somewhat more tentative than for other subtests.

Mazes (WISC-III only)

- Planning ability or foresight.
- Perceptual organization.
- Visual-motor coordination and speed.
- Nonverbal reasoning.

The Mazes subtest is an optional portion of the WISC-III and is not extensively used. Its correlation with the Full Scale IQ is unimpressive (.31), and it is also a poor measure of *g* (9% of its variance may be attributed to *g*). Despite these significant limitations, Mazes can at times provide an additional useful test, particularly with nonverbally oriented children or when a further assessment of planning, sequencing, and perceptual organization is required. Its main advantage is that it is a relatively pure measure of perceptual planning ability.

Individuals with high scores may have an efficient ability to plan ahead and maintain a flexible mental orientation, which further suggests an excellent ability to delay impulsive action (Ireland-Galman, Padilla, & Michael, 1980). Low scores reflect impulsivity and poor visual-motor coordination. Often, unusually low scores may suggest poor reality orientation or organic cerebral impairment, particularly to the frontal areas (Waugh & Bush, 1971).

ASSESSING BRAIN DAMAGE

General Principles

The WAIS-III and WISC-III measure many abilities that are likely to be lowered by brain damage. These include memory, learning, perceptual organization, problem solving, and abstract reasoning. As a result, the Wechsler intelligence scales are typically a core feature of any neuropsychological battery (Groth-Marnat, 2000b; Sullivan & Bowden, 1997). At one time, it was hoped that the Wechsler intelligence scales, along with other more specialized psychological tests, could be used in the actual diagnosis of brain damage. Despite some noteworthy success in this area, it is currently more typical for psychological tests to be used in the assessment of the effects a known lesion is likely to have on a person's cognitive and adaptive functioning. This further highlights the point that the Wechsler intelligence scales, along with other specific tests of neurocognitive ability, are not tests specifically sensitive to brain damage. Rather, they are tests that can reflect the effects of brain damage as well as a variety of other conditions.

During the earlier development of the WAIS and WISC, Wechsler (1958) hoped that brain damage could be discriminated based on relative lowerings in subtests that were most sensitive to neurological impairment. He referred to these brain-sensitive tests as *no-hold* tests (Digit Span, Digit Symbol, Similarities, Block Design) and contrasted them with *hold* tests, which were believed to be far more resistant to impairment (Information, Object Assembly, Picture Completion, Vocabulary). Although the distinction between hold and no-hold tests has some truth, the use of such a distinction in diagnosing brain damage has been found to result in too many misclassifications. Vogt and Heaton (1977) summarized the reasons for this lack of success by pointing out:

- There is no single pattern of brain damage, so it would be expected that highly variable test responses would occur.
- The hold/no-hold distinction does not account for other significant factors, such as the age when the brain damage occurred, environmental variables, education, location of the lesion, and whether the lesion is recent versus chronic.
- Many important abilities related to brain damage still are not measured by the Wechsler intelligence scales.

More recent work supports the theory that there is no specific brain damage profile (Aram & Ekelman, 1986; R. A. Bornstein, 1983; Groth-Marnat et al., 2000; Lezak, 1995; J. Todd, Coolidge, & Satz, 1977). Some persons with brain damage produce low

IQs, whereas for others, IQs are still high. Sometimes, there is a high level of subtest scatter, and, at other times, the scores on the subtests are quite even. Some persons with brain damage produce a high Verbal-Performance split and others do not. This is further complicated because a Verbal-Performance split is more likely to occur for males than for females (R. A. Bornstein & Matarazzo, 1982; A. Kaufman & Lichtenberger, 2002; Lezak, 1995) and for adults but not for children (A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Lezak, 1995). Brain damage may cause a general lowering on all or most subtests and, at other times, there may be a lowering of only specific subtests. The most general indicator for the detection of brain damage is whether a person's scores (either general or specific) are lower than expected given his or her socioeconomic status, age, education, occupation, and other relevant areas of his or her history.

One of the older conventional wisdoms about brain damage is that left hemisphere involvement is more likely to lower the Verbal Scales, whereas right hemisphere involvement results in relatively lower scores on the Performance Scales (see previous discussions under Verbal/Performance IQs, Verbal Scales, and Performance Scales). Reviews of this hypothesis have shown that sometimes this laterality effect has occurred and, at other times, it has not (Aram & Ekelman, 1986; R. A. Bornstein, 1983; A. Kaufman & Lichtenberger, 2002; Larrabee, 1986; Lezak, 1995). On average, right hemisphere lesions produce a V > P discrepancy of 9 points, whereas left hemisphere lesions produce a less marked P > V difference of 4 points (see review by A. Kaufman & Lichtenberger, 2002). Probably the safest approach is that a Verbal-Performance split is not diagnostic of either brain damage in general or, more specifically, damage to one or the other hemisphere. However, a Verbal-Performance split (especially if 15 points or greater) can at times be consistent with this hypothesis. This is especially true if the Verbal-Performance difference is 25 points or greater. More specifically, a lowered Verbal Scale (15 points or greater) suggests the possibility of language impairment. Noteworthy subtests within the Verbal Scales are Arithmetic, Digit Span, and Letter-Number Sequencing (WAIS-III) that, if lowered, suggest difficulties with attending and concentrating. A Performance Scale that is 15 or more points lower than the Verbal Scale suggests impaired perceptual organization abilities. Appropriate caution should be taken to avoid the risk of overinterpreting a person's results and to use further means of investigation, including knowledge of health status, medical history, and additional specialized psychological tests.

Another frequent belief is that brain damage is more likely to lower Performance than Verbal tests. Some good reasons can be given to suggest this may be true. The Performance subtests are timed and, because many persons with brain damage tire easily and have difficulties with concentration and attention, they would be expected to have a particularly difficult time with these tests. Support for this has been found because the Processing Speed Index (Digit Symbol-Coding and Symbol Search) has been lowered with several types of cognitive impairment (D. Fisher, Ledbetter, Cohen, Marmor, & Tulsky, 2000; K. Hawkins, 1998; Psychological Corporation, 1997). From a theoretical perspective, fluid intelligence is tied more to an intact brain structure and also is assessed more clearly by the ongoing problem-solving tasks presented in the Performance subtests. Thus, a destruction of brain tissue would be more likely to lower fluid intelligence, which would be reflected in lowered Performance subtest scores.

This hypothesis can be further assessed by calculating Horn's WAIS-III or WISC-III subtest groupings for fluid intelligence (see "WAIS-III/WISC-III Successive Level Interpretation Procedure" section, Level II, Step c). Although there is some basis for accepting the preceding assumptions, there are also many exceptions. Russell (1979) and Zilmer, Waechtler, Harris, Khan, and Fowler (1992) found that left hemisphere damage caused a lowering in both WAIS/WAIS-R Performance and Verbal subtests, whereas right hemisphere and diffuse damage resulted in the expected lowering, primarily in Performance subtests.

A. Kaufman and Lichtenberger (2002) suggest that an important reason for this relatively small V > P effect for unilateral left lesion patients is that different hemispheres do not so much process different types of information (verbal content versus visual-spatial content), but more that the left hemisphere processes information sequentially whereas the right hemisphere processes information simultaneously (see Springer & Deutsch, 1998). This is supported by the observation that adult left-lesion patients do worst on Arithmetic, Digit Span, and Digit Symbol-Coding, all of which require sequencing (and comprise the WAIS-III/WISC-III Working Memory/Freedom from Distractibility factor). The WAIS-R difference between unilateral left lesion patients' average subtest scores on Perceptual Organization (8.7) and Freedom from Distractibility (6.8) is nearly 2 subscale points. Thus, it might be more useful to assess the relative extent of lowering on unilateral left lesion patients' Freedom from Distractibility than to merely assess the extent of their P > V difference. Future research on the WAIS-III's Working Memory Index (Arithmetic, Digit Span, Letter-Number Sequencing) would also be likely to support these findings.

Many of the inferences related to brain damage depend on profile analysis. Useful material relevant to brain damage can be found in the discussion of Levels II through V under the "Interpretation Procedure" section in this chapter and in the relevant discussions for each subtest in the "Wechsler Subtests" section of this chapter. Much of this interpretation depends on hypothesis testing in which the practitioner integrates knowledge about the person, brain function, Wechsler subtests, and past clinical experience. Often, no clear, empirically based guidelines exist. Accuracy of any inferences are based partially on whether they make neuropsychological sense. However, one generally accepted principle is that intersubtest scatter is most likely to occur with focal lesions of recent origin (A. Kaufman & Lichtenberger, 2002). In contrast, general lowering of all abilities (low subtest scatter) is more likely with either chronic lesions or with diffuse degenerating diseases (e.g., exposure to neurotoxins; Groth-Marnat, 1993; L. Miller, 1993).

One useful strategy developed by Kaplan and her colleagues is to work toward parceling out the underlying processes responsible for scores on the Wechsler intelligence scales (Milberg et al., 1996). Alternative administration guidelines, error categories, useful tables, and interpretive procedures have been developed for both the WAIS-R (E. Kaplan et al., 1991; with plans for the WAIS-III) and WISC-III (E. Kaplan et al., 1999). For example, a clinician might be interested to know if a client's poor performance on Information or Vocabulary resulted from lack of knowledge or problems with retrieval. This might be determined by presenting him or her with multiple-choice formats that assist (recognition of correct answers) them with the retrieval process. If a client does significantly better on the multiple-choice format than the standard format,

it suggests that the lowering was caused by retrieval difficulties. The new WAIS-III Digit Symbol-Coding optional procedures (Incidental Learning and Digit Symbol-Copy) were originally derived from Kaplan et al.'s (1991) WAIS-R as a Neuropsychological Instrument (WAIS-R NI) and, as discussed previously, can assist in determining if a poor performance resulted more from poor memory or graphomotor (psychomotor) slowing. Another strategy built in to the process approach is to carefully investigate various error categories (Groth-Marnat et al., 2000; E. Kaplan et al., 1991, 1999). For example, visual neglect might be indicated by not noticing details on the left (usually) side of pictures on Picture Completion or making errors on the left side of the designs for Block Design.

When the preceding strategies, principles, and cautions are taken into account, clinicians can generate and test useful hypotheses developed from different patterns of subtest scores. The following list summarizes some of the most frequently supported hypotheses about specific subtests or patterns of subtests:

- Digit Symbol-Coding is the most brain-sensitive Wechsler subtest and can be lowered by lesions in any location. A lowering implies difficulties with speed of information processing and/or learning, sequencing, rote learning, concentration (especially with lowerings in Digit Span and Arithmetic), visual-motor abilities, and speed of processing or learning (Lezak, 1995; Reitan & Wolfson, 1992). The WAIS-III combination of Digit Symbol-Coding and Symbol Search (Processing Speed Index) has been found to be the most frequently lowered group of subtests among a wide variety of brain-impaired populations (K. Hawkins, 1998; Psychological Corporation, 1997).
- Block Design is also brain sensitive, especially to either left or right parietal lesions (Golden, 1979; Lezak, 1995; McFie, 1960, 1969). A lowering implies visual-spatial problems (especially combined with a lowering in Object Assembly) and possible difficulty in constructing objects (constructional apraxia: note quality of drawings; J. H. Kramer et al., 1999; Zilmer, Bell, Fowler, Newman, & Stutts, 1991).
- Picture Arrangement lowering is consistent with right anterior temporal and possibly right frontal lesions (Reitan, 1974b; Reitan & Wolfson, 1993; Russell, 1979). In some cases, Picture Arrangement might also be lowered by left hemisphere lesions if there is a resulting impairment in following directions and/or conceptual skills.
- Both Digit Span and Arithmetic are frequently lowered in brain-damaged populations, particularly with left hemisphere lesions (A. Kaufman & Lichtenberger, 2002; Lezak, 1995; McFie, 1960, 1969). Lowering suggests poor concentration and attention and, if Digits Backward is significantly lower than Digits Forward (generally 5 or more digits), a significantly reduced level of mental flexibility and/or difficulty forming and maintaining a visual image of the digits. It may also suggest difficulties in a person's executive functions related to selecting a key stimulus, attending to it, and maintaining the information in short-term storage, while simultaneously performing other mental tasks (Wielkiewicz, 1990).
- Processing Speed (composed of Symbol Search and Digit Symbol-Coding) is the subtest that is most sensitive to the impact of most forms of cognitive impairment (D. Fisher et al., 2000; K. Hawkins, 1998; Psychological Corporation, 1997.)

- Vocabulary, Information, and Picture Completion have often been used as a rough estimate of a person's premorbid level of functioning because they are relatively unaffected by lesions. An important exception is that children who are brain damaged often score lowest on the Vocabulary subtest (Boll, 1974; Reitan, 1974b; Reitan & Wolfson, 1992). In addition, Information and Vocabulary are generally lowered (especially relative to Similarities) in patients with left temporal damage, suggesting difficulties with word comprehension, retrieval, and language expression (Dobbins & Russell, 1990). Another hold test, Picture Completion, while usually resistant to brain damage, might be lowered because of difficulties involving vision, especially visual agnosia (difficulty recognizing objects; E. Kaplan et al., 1991, 1999). Thus, always considering Vocabulary, Information, and Picture Completion as indicators of premorbid functioning can potentially result in incorrect inferences and should be interpreted in relation to what is known about brain-behavior relationships.
- The Similarities subtest, especially in relation to Information and Vocabulary, is most likely to be lowered with left frontal lesions and suggests difficulty with verbal reasoning and verbal concept formation (Dobbins & Russell, 1990).
- Qualitative responses, particularly related to error categories (even when the subtests are not lowered), can provide useful information related to brain damage. Some responses might suggest poor judgment and impulsivity, whereas others might indicate concrete thinking in which the person is bound by the stimulus value of the item (e.g., winter defined as "wet, cold" rather than the more abstract reference to a season; or the clang response that "ponder" means "to pound"). Other persons might report they once knew the answer but have forgotten, which can be assessed through WAIS-R NI/WISC-III PI multiple-choice options. Diffuse brain damage (but not focal) might also be consistent with a high degree of intratest scatter in which the client misses easy items but correctly answers later, more difficult ones (Mittenberg et al., 1989). This suggests retrieval failure and/or the random loss of previously stored information. This intrasubtest scatter is most likely to occur on Vocabulary, Comprehension, Information, Similarities, and Picture Completion.

Estimating Premorbid IQ

Neuropsychologists are frequently confronted with the need to estimate a client's premorbid level of functioning. In an ideal situation, previous IQ results derived before the injury could be obtained and compared with his or her current level of functioning. Even in this situation, clinicians should be aware that a decline in *overall* performance should not be inferred unless there is a significantly lower current IQ than had been obtained from a premorbid IQ assessment. A discrepancy of 12 or more WAIS-R Full Scale IQ points would result in an 80% accurate detection of adults (WAIS-III) who had actually suffered a cognitive decline (Graves, Carswell, & Snow, 1999). It should also be stressed that there still might be quite specific areas of decline that are not sensitive to the global measure of IQ scores.

In most cases, premorbid IQ results are not available; therefore, clinicians must rely on other strategies to infer premorbid ability. These strategies include historical achievement-based records, current measures of ability that are not sensitive to decline

("hold" measures), demographic-based regression equations, or a combination of these. Useful historical records might include grade point average, SAT scores, work achievement records, achievement tests, or peer ratings. The age of the person, as well as relevant aspects of the injury (i.e., size and location of the lesion, recency of injury), might also be important to consider.

A further strategy for estimating premorbid ability is to note performances on Wechsler subtests that are considered most resistant to neurological impairment (Information, Picture Completion, and especially Vocabulary). As discussed previously, these subtests have often been considered to reflect the person's past level of functioning and are, therefore, referred to as *hold* subtests. Administering an achievement test such as the Wide Range Achievement Test (WRAT-III) or Wechsler Individual Achievement Test (WIAT) might also accomplish a similar purpose. One difficulty is that for many clients, especially those who are well educated, this method is likely to overestimate premorbid IQ. In contrast, it would be likely to underestimate premorbid IQ for subgroups whose premorbid Performance Scales are typically greater than Verbal Scales (i.e., Native Americans, Hispanics, bilinguals, persons with low educational attainment, blue-collar workers).

A related technique is to consider the person's two or three highest subtests (regardless of whether the subtests are brain-sensitive or non-brain-sensitive) and then use these to estimate the person's premorbid level of functioning. Despite its occasional usefulness, this procedure is likely to result in a high number of misclassifications because it does not consider crucial factors such as the person's age, educational level, or location of the lesion (Matarazzo & Prifitera, 1989).

A variation of this hold procedure is to use a reading test such as the National Adult Reading Test (NART; H. Nelson & Williams, 1991) or Wechsler Test of Adult Reading (WTAR; Wechsler, 2001). The NART and WTAR were designed by selecting 50 irregularly spelled words (i.e., yacht, naive) that are unlikely to be pronounced correctly unless the client has previous knowledge of the words. This relatively pure recognition task places minimal demands on problem-solving abilities. A NART-estimated WAIS-R Full Scale IQ 20 points higher than a person's obtained IQ suggests intellectual decline (80% accuracy for those with actual decline; Graves et al., 1999). However, this assumes that the injury would not have affected the person's reading ability. The WTAR has the advantage that it has been co-normed with the WAIS-III and WMS-III. Despite their usefulness, the previous caveats related to demographics (ethnicity, education) would also be relevant for reading tests such as the NART/NART-R and WTAR.

Other efforts to determine premorbid IQ have used regression equations based on demographic variables (education, occupation, etc.). One of the most extensively researched is the Barona Index (Barona, Reynolds, & Chastain, 1984). To correctly classify (80% accuracy) clients with true cognitive decline, a discrepancy of 25 IQ points would be required (Graves et al., 1999). Unfortunately, this discrepancy is sufficiently large such that other more straightforward procedures (i.e., previous work performance, grade point average, medical records) would be likely to be more accurate. In addition, the index is likely to be inaccurate for persons with either extremely high (above 120) or extremely low (below 69) IQs (Barona et al., 1984; Graves et al., 1999; Veiel & Kooperman, 2001), and the formulas are likely to overestimate most premorbid IQ levels (Eppinger, Craig, Adams, & Parsons, 1987).

A final strategy is to combine various measures such as the NART and demographics or performance on specific subtests with demographics. Such procedures have generally resulted in slight incremental increases beyond the NART or demographics alone (Grave et al., 1999; Vanderploeg, Schinka, & Axelrod, 1996). Vanderploeg et al. found that the best predictor of WAIS-R Full Scale IQ could be made by calculating the following three regression equations and using the one that resulted in the highest IQ estimate (BEST 3 approach):

WAIS-R FSIQ = 3.55 (Information) + 1.00 (SES) + 58.70

WAIS-R FSIQ = 3.78 (Vocabulary) + 0.70 (SES) + 59.09

WAIS-R FSIQ = 2.94 (Picture Completion) + 2.13 (SES) + 1.62 (Age) + 49.41

These calculations can be made by inserting the following variable codes:

Age: 16–17 years = 1; 18–19 = 2; 20–24 = 3; 25–34 = 4; 35–44 = 5; 45–54 = 6; 55–64 = 7; 65-69 = 8; 70–74 = 9

Education: 0–7 years = 1; 8 = 2; 9–11 = 3; 12 = 4; 13–15 = 5; 16+ = 6

Occupation: Unemployed = 1; farm laborers, farm foreman, and laborers (unskilled) = 2; operatives, service workers, farmers, and farm managers (semiskilled) = 3; craftsmen and foremen (skilled workers) = 4; managers, officials, proprietors, clerical, and sales workers = 5; professional and technical = 6

SES: Sum of education code and occupation code (If unemployed, SES = 2 × Education)

The correlation with the actual Full Scale IQ is .84, and the standard error of estimate was 9.10 using the equation with Information, 8.64 for Vocabulary, and 9.57 for Picture Completion. To infer overall cognitive decline, discrepancies of 18 points or more should be documented. This is clearly superior to the estimated 25-point discrepancy required for the Barona index. However, these formulas were calculated using the WAIS-R. Because WAIS-III Full Scale IQ scores are, on average, 3 points higher than scores derived from the WAIS-R (and, therefore, estimated by these equations), an additional 3 points (21 points in total) would be needed to infer cognitive decline if current IQ scores were obtained with the WAIS-III. In addition, the BEST 3 approach tends to slightly overpredict scores at the low IQ range but underpredict estimates in the high IQ range.

In contrast to adult BEST-3 premorbid estimates, research with children has found that an equation based on demographics alone is equally as effective in differentiating people with brain damage from non-brain-damaged persons as equations using a combination of demographics and WISC-III subtests (Vanderploeg, Schinka, Baum, Tremont, & Mittenberg, 1998). Thus, the following formula based on demographics alone is recommended:

FSIQ = 5.44 (Mean parental education) + 2.80 (White/non-White)
− 9.01 (Black/non-Black) + 81.68

This equation can be calculated by inserting the following variable codes:

Mean parental education: 0–8 years = 1; 9–11 = 2; 12 years (or GED) = 3; 13–15 years = 4; 16+ = 5

Ethnicity: Two coded variables: White/non-White (White = 1; non-White = 0) and Black/non-Black (Black = 1; non-Black = 0). Hispanics would be uniquely coded as 0 on both White/non-White and Black/non-Black (the regression equation should not be used for ethnic groups other than White, Black, or Hispanic).

However, when using a discrepancy cutoff of 13, only 64% of people with brain damage were correctly classified and 89% of normal controls were correctly classified (Vanderploeg et al., 1998).

As would be expected, estimating premorbid IQ has been a controversial procedure, particularly in a forensic context (see Veiel & Koopman, 2001). The following review points seem crucial. First, the previous equations should be used to supplement but not replace a careful evaluation of crucial information such as work history and medical records. In addition, formal cutoffs should be used. Rarely, for example, would an obtained IQ 5 to 10 points below the estimated "premorbid IQ" suggest actual cognitive decline in a person's *overall* ability. However, this still does not preclude the possible presence of quite specific deficits (i.e., facial recognition, short-term visual memory). The likelihood of errors increases when equations based on demographics or subtests are used with persons with IQs suspected of being extremely high or extremely low (below 80 or above 120).

Alzheimer's Disease

The initial symptoms of Alzheimer's disease are characterized by apathy, a decline in short-term memory, and difficulties with problem solving. Underlying these changes are reductions in cholinergic activity. Currently, neuropsychological assessment, particularly with the Wechsler intelligence scales, is one of a variety of diagnostic procedures to enhance diagnosis. Nonverbal abilities seem to be more sensitive to impairment than verbal abilities. Earlier research with the WAIS-R found that a full 52% of Alzheimer's disease patients had Verbal greater than Performance scores of 15 points or more (Fuld, 1984). Similarly, WAIS-III Verbal scores have been found to be 10 points higher than Performance subtests for a group of patients with "probable" Alzheimer's (Psychological Corporation, 1997). The lowest index scores were for Processing Speed (mean = 79.6) with some lowerings in Perceptual Organization (mean = 84.8) and Working Memory (mean = 87.2).

A specific WAIS-R Alzheimer's profile developed by Fuld (Fuld, 1983, 1984) found that Information and Vocabulary were relatively higher than Similarities and Block Design, and Digit Symbol and Block Design were lowest. This pattern makes conceptual sense in that Information and Vocabulary are relatively resistant to deterioration, reflect crystallized abilities, and are correspondingly the highest subtests in the profile. In contrast, Digit Symbol and Block Design are relatively sensitive to deterioration, reflect areas of fluid intelligence and, along with Object Assembly, are the

lowest subtests in the profile. An extensive review of the Fuld profile using 18 studies concluded that sensitivity (proportion of true positives) to Alzheimer's disease was a very low 24.1% (Massman & Bigler, 1993). In contrast, the profile's specificity (proportion of true negatives) was 93.3%. This means that more accurate diagnoses are likely to be achieved through using the WAIS-III in combination with specific measures of memory (i.e., WMS-III) or specialized dementia batteries (i.e., CERAD battery). In addition, research on the Fuld or similar profiles needs to be conducted with the WAIS-III.

ASSESSING ADDITIONAL SPECIAL POPULATIONS

Learning Disabilities

Learning disabilities make up a complex, heterogeneous, loosely defined disorder with a wide variety of manifestations and many different theories regarding causation (A. Kaufman & Kaufman, 2002; Sattler, 2002; L. Siegel, 1999). A central component of all definitions is that learning disabilities involve difficulties in developing skills in reading (most commonly), writing, listening, speaking, reasoning, spelling, or math. This is sometimes summarized as poor information processing. Further essential features are these: Learning-disabled persons have adequate intelligence, show a significant discrepancy between achievement and intellectual ability, and have a disorder that is considered primarily intrinsic to the person, presumably because of central nervous system dysfunction. The underachievement cannot be primarily the result of an intellectual disability (mental retardation), brain damage, behavior problems, sensory handicaps, or environmental disadvantage.

The major purpose of learning disability assessment is to identify a client's strengths and weaknesses to be able to decide on an appropriate placement and to design an optimal program. Relevant areas to assess include developmental-cognitive processes, achievement, environmental demands, reactions of others to the client's difficulties, and the possible interaction of additional factors, such as fear of failure, overall level of interpersonal adjustment, and family history of similar difficulties. The Wechsler scales are typically considered essential as a means of identifying the client's overall level of functioning and specific cognitive strengths and weaknesses and to eliminate the possibility of intellectual disability (mental retardation). Other tests are usually required; for example, achievement tests, measures of adaptive behavior, visual-motor tests, assessments of auditory and visual processing, and measures of emotional and behavioral problems (see L. Siegel, 1999).

Considerable effort has been placed into searching for a specific Wechsler scale profile that is unique to learning-disabled populations (see Level IIIb in "Interpretation Procedure" section). There is some evidence for a WAIS-III ACID profile (Arithmetic, Coding/Digit Symbol, Information, and Digit Span) in that 24% of those diagnosed with learning disabilities had a partial (three out of the four subtests as the lowest scores) ACID profile and 6.5% had a full (all four of the subtests as the lowest) ACID profile (Psychological Corporation, 1997). This is higher than the standardization sample. The WAIS-III index scores of Working Memory and Processing Speed

(compared to Perceptual Organization and Verbal Comprehension) were also found to be particularly low among a sample of adults diagnosed with reading disabilities (Psychological Corporation, 1997). This has led A. Kaufman and Lichtenberger (1999, 2002) to suggest the possible utility of combining the five subtests in these lowest indexes into a SCALD profile (Symbol Search, Digit Symbol-Coding, Arithmetic, Letter-Number Sequencing, Digit Span). The ACID profile has also received some support with the WISC-III in that most studies have found that approximately 20% of persons with learning disabilities had either a partial or full ACID profile (Mayes, Calhoun, & Crowell, 1998; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Stanton & Reynolds, 1998).

A somewhat similar WISC-III profile substitutes the new Symbol Search subtest for Information, resulting in the SCAD (Symbol Search, Coding, Arithmetic, Digit Span) profile. These four subtests emphasize the functions of speed of information processing, visual short-term memory, and visual-motor coordination (Symbol Search and Coding), as well as number ability and sequencing (Arithmetic and Digit Span). These are specifically the types of functions that many learning-disabled individuals (as well as many other types of persons with brain dysfunctions) have difficulty with. Accordingly, children with learning disabilities and attention deficit disorder have been found to score particularly low on the SCAD profile (A. Kaufman, 1994; Mayes et al., 1998; Stanton & Reynolds, 1998). Similarly, children diagnosed with ADHD have performed relatively poorly on the WISC-III Freedom from Distractibility factor (Anastopoulos, Spisto, & Maher, 1994). This finding should be used with caution, however, because a relatively large proportion of children with ADHD still do not have this profile. In addition, S. Ward, Ward, Hatt, Young, and Mollner (1995) did not find support for the SCAD profile among learning-disabled children.

A further approach to understanding learning disabilities and related disorders is using Bannatyne's categories, which conceptualize learning-disabled performances as highest on subtests requiring spatial abilities (Object Assembly, Block Design, Picture Completion) in which little or no sequencing is required (Bannatyne, 1974). Conceptual skills are intermediate (Comprehension, Similarities, Vocabulary), and subtests requiring sequencing abilities (Digit Span, Digit Symbol-Coding, Picture Arrangement) are lowest. Thus, their spatial abilities are believed to be greater than their conceptual abilities, which, in turn, are greater than their sequential abilities. A fourth category, Acquired Knowledge (Information, Arithmetic, Vocabulary) is also sometimes used as a rough index of the extent to which the person has accumulated school-related facts and skills (see Level IIIc of "Interpretation Procedures" section). Even though these findings might suggest a greater degree of subtest scatter among learning-disabled persons, this has not been supported by research (Greenway & Milne, 1999).

Collectively, the preceding profiles suggest that many learning-disabled individuals perform best on tasks requiring holistic, right brain, simultaneous processing (Object Assembly, Picture Completion, Block Design) and worst on those requiring sequential processing (Digit Span, Digit Symbol/Coding, Picture Arrangement), which is expressed in difficulties with planning, reading, and numerical ability. Wielkiewicz (1990) has further suggested that these subtests indicate a poorly functioning executive ability in which the individual experiences difficulty attending to stimuli while simultaneously performing other mental tasks.

Reviews and cross-validation of Bannatyne's and ACID/SCAD profiles have produced inconsistent results (see Groth-Marnat, 2002). Only some groups of learning-disabled students in some studies showed the Bannatyne Spatial > Conceptual > Sequential pattern (Katz et al., 1993; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002). This is not surprising given the many different modes of expression found under the umbrella term of "learning disabilities" (A. Kaufman & Kaufman, 2002). In addition, Bannatyne's pattern has not been found to be unique to learning disabilities, but frequently occurs in a diverse number of groups including juvenile delinquents and emotionally handicapped children (see Groth-Marnat, 2002). Although only minimal support exists for Bannatyne's categories as a diagnosis for learning disabilities, they are far from useless. The four categories (Spatial, Conceptual, Sequential, Acquired Knowledge) can be invaluable for interpreting relative strengths and weaknesses for learning-disabled persons as well as for other groups. While research has not been able to produce a unique "learning-disabled profile," the research invested in this effort has resulted in a useful means of analyzing Wechsler scale profiles.

Given the previous research, the following conclusions are warranted (adapted from Groth-Marnat, 2002):

- The Full Scale IQ can be most appropriately used in the assessment of persons with learning disabilities to estimate their overall potential and assist in excluding possible explanations for poor academic performance, other than learning disabilities (i.e., intellectual disabilities/mental retardation).
- There is moderate-to-equivocal evidence that some profiles (relatively low Processing Speed and Working Memory/Freedom from Distractibility, Spatial > Conceptual > Sequential, ACID, SCAD, SCALD) occur more frequently in learning-disabled populations compared to the general population.
- These profiles are not unique to learning disabilities but often occur in other groups as well (juvenile delinquents, ADHD, emotionally handicapped).
- If a person does have a "learning-disabled" Wechsler profile (ACID, etc.), it is *consistent with, although not necessarily diagnostic of,* learning disabilities.
- The majority of learning-disabled persons *do not* have Wechsler "learning disabled" profiles. Thus, the absence of one of the profiles *does not exclude* a diagnosis of learning disabilities.
- The various patterns of Wechsler subtests can, at times, be used to further understand individual cases of persons experiencing learning difficulties.

Mental Retardation (Intellectual Disability)

Mental retardation (intellectual disability) is a nonspecific, heterogeneous disorder that occurs during a person's early developmental stages (birth to 18 years; J. Jacobson & Mulick, 1996). It is defined in part as involving subaverage general intellectual performance, which in turn is defined as less than 2 standard deviations below average. Of equal importance are difficulties in adaptive behavior, and any assessment of intellectual disability must demonstrate both a low intelligence level (2 standard deviations below the mean) and evidence that the person cannot function independently or deal effectively with day-to-day life problems (American Psychiatric Association, 1994).

This must include at least two adaptive skill areas including communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure, and work (J. Jacobson & Mulick, 1996). Classification of mental retardation (intellectual disabilities) should identify the person's psychological and emotional strengths and weaknesses, overall physical health, and current environmental placement. The American Association of Mental Retardation (AAMR) guidelines (see J. Jacobson & Mulick, 1996) stress that this should lead to a profile that places less emphasis on describing the level of disability (mild, moderate, severe) and more on identifying the types and intensities of supports required by the person. These might be intermittent, limited, extensive, or pervasive. Thus, there has been a recent move away from describing the disability in favor of using information about the person to identify how the person's functioning could best be optimized by the best support available for the person. With appropriate supports, the person's functioning should be able to improve over time. In addition, assessment should take into consideration cultural and linguistic diversity, the context of the community environment, and balance out the individual's adaptive limitations with his or her adaptive skills and personal capabilities (see Schalock et al., 1994).

The AAMR guidelines emphasize the interaction of the person with the environment and, in particular, they encourage any assessment to focus on the level and intensity of required support with a philosophy of empowering the person. As such, there has been a relative deemphasis on the global IQ score, along with the elimination of person-oriented levels of disability. This does not mean that IQ scores are not important, but there is more of a focus on treatment and community-oriented descriptions. In somewhat of a contrast to this trend are the guidelines in the *DSM-IV* (1994), which continue to classify the degree of severity (and corresponding diagnostic code) based on the following IQ ranges: 50–55 to 70 (mild), 35–40 to 50–55 (moderate), 20–25 to 35–40 (severe), below 20–25 (profound) and severity unspecified (mental retardation is presumed to exist but intelligence is untestable by standard tests). The implications are that, for most contexts, clinicians should follow the AAMR guidelines because they are more useful, more clearly tied to recommendations, represent the most current thinking in the field, and are in accordance with national recommendations. However, there may be certain situations in some contexts where *DSM-IV* guidelines might be required.

Although mental retardation (intellectual disability) is a heterogeneous disorder, there is consensus that it consists of two general categories. Nonorganic (or familial) retardation is caused by low genetic inheritance, poor environment, and possibly some organic factors. Persons with familial retardation constitute the upper realms of intelligence (50 to 69) and adaptive functioning among persons with intellectual disabilities and can be educated. Organic retardation is frequently severe (IQ less than 50) and is more closely associated with neurological dysfunction and genetic impairment. Persons with this disorder typically require more supervision and care but are typically able to be taught to manage some routine day-to-day activities.

A typical assessment battery for the diagnosis and assessment of mental retardation (intellectual disability) includes the WISC-III or other individually administered intelligence tests (K-ABC, Stanford-Binet), an achievement test (Wide Range Achievement Test-III, Wechsler Individual Achievement Test, Kaufman Test of Educational Achievement), and measures of adaptive functioning (Adaptive Behavior Assessment System, AAMD Adaptive Behavior Scale, or Vineland Adaptive Behavior Scales).

Further information from interviews, behavioral observations, and medical records are also essential. An important purpose of a test such as the WISC-III is to establish the client's general intelligence level so that it can be placed into the context of other relevant information. The AAMR guidelines point out that, when determining the cutoff IQ for diagnosis, the range of error of the test should be taken into account. This means that the IQ cutoff criteria are somewhere between 70 and 75. The most difficult subtests for mentally retarded persons are Information, Arithmetic, and Vocabulary (primarily the Verbal Comprehension factor), while the easiest subtests are Picture Completion and Object Assembly (primarily the Perceptual Organization factor; Mueller, Dash, Matheson, & Short, 1984).

Gifted Children

Gifted children are frequently defined as having Verbal or Performance IQs of 130 or higher. Children who have a single outstanding ability such as art, music, or math are also frequently classified as gifted even though their IQs may not necessarily be above 130. A further caution is that the WISC-III places considerable emphasis on speeded performance. Thus, a person who was generally gifted, but did not express this giftedness in a rapid manner, may not do particularly well on the WISC-III. Although the WISC-III might be frequently used to identify giftedness, the Stanford-Binet may be somewhat more effective because it has a higher ceiling than the WISC-III. However, neither may be particularly good if a single outstanding ability is used to determine whether a particular child is gifted. Additional assessment strategies for children should include samples of their work, achievement tests, rating forms, or designation by a highly qualified person.

An essential goal of assessing for giftedness is to optimize (rather than "normalize") the child's abilities so that a greater likelihood exists that the child will eventually make a significant contribution to society. This implies that the assessment can recommend an appropriate educational placement and provide general guidelines for program planning. IQ, in itself, is in many ways a limited definition of giftedness. Many persons with extremely high IQs do not accomplish anything of significance. A high IQ (or outstanding talent in a specific area) is merely one of a variety of prerequisites. The interactions of internal motivation, discipline, and environmental opportunities, such as appropriate instruction, are of equal importance.

Caution should also be used with tests such as the WISC-III to assess gifted persons who demonstrate high creativity. Often, highly intelligent people are not particularly creative, which is supported by the low correlation between intelligence tests and creativity (Amabile, 1983). For abilities such as artistic or musical creativity, measures outside IQ testing may prove to be of greater importance. These might include a list of creative achievements, nomination by a qualified person, and specific tests of creativity.

Ethnic Minorities

Intelligence tests have frequently been criticized for being limited in assessing ethnic minorities. A detailed discussion of this issue is included in Chapter 2 (see "Test Bias and Use with Minority Groups" section). However, several additional guidelines

should be noted. Often, it is essential to be familiar with the values and beliefs of the client's culture as well as relevant research. This is underscored by the observation that the degree of cultural difference between an interviewer and client has been found to be related to the amount of inaccurate perceptions (Malpass & Kravitz, 1969; P. Shapiro & Penrod, 1986). A clinician should determine the language most familiar to the client and establish the extent and manner in which any language difference might bias the test results. Of related and equal importance is the degree to which clients have assimilated into the dominant culture. Directions and pronunciation should be particularly clear. The examiner also needs to pay particular attention to the importance of rapport and motivation.

Probably the most important strategy is to maintain a flexible attitude, combined with the use of alternative assessment strategies. This strategy might include a variety of nonverbal techniques, such as the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998), Raven's Progressive Matrices Test, or emphasis on the Performance Scales of the WAIS-III/WISC-III. In addition, dynamic testing shows promise in assessing the extent to which a client can benefit from various ongoing learning opportunities (learning potential; Grigorenko & Sternberg, 1998). Material beyond merely tests should also have a greater significance (teacher reports, discussions with parents, history, behavioral observations).

SHORT FORMS

Dozens of short forms for the Wechsler intelligence scales have been developed to provide a more time-efficient means of estimating IQ. These short forms reduce administration time by either giving selected subtests or deleting specific items (early easy ones, odd or even items). Although time-efficient, these short forms tend to provide less information about a person's cognitive abilities, produce a wider band of error than a full administration, result in less clinical information, and are often of questionable accuracy when used for intelligence classifications (Silverstein, 1990). However, short forms can serve appropriately as screening devices, which are best used when the purpose of evaluation is other than for intellectual assessment. The results can be used either as a rough indicator of intelligence, or as a basis for determining whether a more complete cognitive assessment is necessary. None of the short forms should be confused with a full intellectual assessment or even with a valid indicator of IQ (J. Ryan & Ward, 1999). For this reason, it is important to clearly specify on the report that the indicated IQ is an estimate (indicate as *Est* next to the IQ score) and that a "brief WAIS-III/WISC-III" was given. If this is not specified, the IQ derived from the short form may be confused with a full administration and later decisions may be incorrectly based on the misleadingly described results.

The basic requirement for any short form is a minimum correlation of .90 with the full administration. Even at the .90 level, the band of error is considerably wider than for an IQ derived from a full administration. Calculations indicate that at a .90 correlation, two thirds of the IQs fall within 9 points of a person's actual IQ and a full one third are 10 or more points away from the actual IQ (L. Schwartz & Levitt, 1960). In addition to these psychometric considerations, short forms might be selected based on

the type of clinical information needed, or special client characteristics (i.e., handicapped, non-English-speaking background).

Many clinicians calculate short form IQs by prorating the subtest scores—by calculating the mean subtest score for the subtests that were given. This mean can then be multiplied by the total number of Performance, Verbal, or Full Scale subtests to derive the equivalent of the Verbal, Performance, and/or Full Scale sum of scaled scores. Once this estimate of sum of scaled scores has been determined, relevant tables in the manual(s) can be consulted to determine the estimated IQs. Unfortunately, prorating may produce error by failing to consider the relative reliabilities of the different subtests that were used. To counter this, the WAIS-III manual allows examiners to prorate sums of scaled scores based on five Verbal and four Performance subtests (see Table A.10 in the *WAIS-III Scoring and Administration Manual*). In addition, Sattler (2001) has provided a formula (see Sattler, 2001, pp. 256–257, Exhibit 8–4) for obtaining deviation IQs from short forms. He has also provided tables for converting scores on various combinations of short forms into IQs (see pp. 828–835 for the WAIS-III and pp. 774–782 for the WISC-III).

Wechsler Abbreviated Measure of Intelligence (WASI)

The Psychological Corporation developed the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1997) as a means of providing clinicians and researchers with a short, reliable measure of intelligence linked to the WAIS-III (and WISC-III). The WASI includes four subtests (Vocabulary, Similarities, Block Design, and Matrix Reasoning), which have a similar format and similar content as the WAIS-III subtests with the same names. The selection of these subtests was based in part on high loadings on *g,* along with evidence suggesting bilateral hemispheric activation on most complex cognitive tasks (Springer & Deutsch, 1998). The WASI yields both Verbal and Performance IQs, as well as a Full Scale IQ. The WASI was nationally standardized using a population ranging between ages 6 and 89. Because the subtests were linked to the longer Wechsler intelligence scales, the WASI provides reliable estimates of full WAIS-III and WISC-III IQs. Administration time can be reduced even further by using a two-subtest form (Vocabulary and Matrix Reasoning), which takes approximately 15 minutes but yields only a Full Scale IQ estimate.

Best Two- and Three-Subtest Short Forms

One of the most frequently used two-subtest WAIS-III/WISC-III short forms uses Vocabulary and Block Design. Administration time is approximately 20 minutes and correlations with the full-administration Full Scale IQ are generally in the .90 range (Sattler, 2001). In two-thirds of the cases, IQs fall within 7 points of a person's actual IQ, and one-third of the scores have an error of eight points or greater. Conceptually, Vocabulary and Block Design are good tests to use because they are both good measures of *g,* are quite stable, and represent a sample subtest from both the Performance and Verbal scales. However, research with the WAIS-R suggests it might potentially underestimate the IQs of African Americans because these two subtests are typically their lowest scores (A. Kaufman et al., 1988). Furthermore, persons with high IQs are

likely to have a greater margin of error when short forms are used to estimate their IQs because of the greater degree of subtest scatter among this subgroup (Matarazzo, Daniel, Prifitera, & Herman, 1988). If examiners wish to add a third subtest, the inclusion of Similarities, Information, Comprehension, Picture Arrangement, and Picture Completion have each been found to increase correlations into the low .90s (Sattler, 2001). An "amazingly short" form made up of the very short administration time subtests of Information and Picture Completion (WISC-III conversion to standard score = 2.9 (I + PC) + 42; A. Kaufman, Kaufman, Ramaswamy, & McLean, 1996) has been found to have correlations in the mid .80s to low .90s (Sattler, 2001).

Best Four-Subtest Short Forms

A possible four subtest combination includes Vocabulary, Arithmetic, Block Design, and Picture Arrangement. Correlations with the Full Scale IQ range from .93 to .95 for the WAIS-III and WISC-III (Sattler, 2001). Research with the WAIS-R indicated that these four subtests are usually excellent in detecting abnormal cognitive functioning (J. Ryan, Georgemiller, & McKinney, 1984). The inclusion of Arithmetic with Vocabulary and Block Design provides an assessment of auditory attention, along with an important indicator of how effectively the person functions in the real world. Picture Arrangement provides information on a person's knowledge of sequencing and his or her relative perceptiveness about common social situations. An important caution is that any short-form combination of Vocabulary, Block Design, Arithmetic, or Picture Arrangement is likely to overestimate the IQs of patients referred for neuropsychological evaluation (Roth, Hughes, Mankowski, & Crosson, 1984). Additional short forms using any four combinations of Vocabulary, Block Design, Arithmetic, Matrix Reasoning, Picture Arrangement, Information, Comprehension, Similarities, or Picture Completion are also likely to produce correlations with the Full Scale IQ in the low to mid-.90s (Sattler, 2001).

A. Kaufman et al. (1996) evaluated WISC-III four-subtest short forms based on clinical, practical, and psychometric considerations and recommended that the overall best tetrad was composed of Similarities-Arithmetic-Picture Completion-Block Design. Total administration time was approximately 27 minutes, scoring time was relatively brief, and it was found to be as psychometrically sound as other combinations. Conversion formulas to estimate Full Scale IQs use the sum of the four scaled scores (S + A + 1 + PC + 1 + BD; abbreviated as simply Xc) but vary according to the following age groups: ages 6, 8 to 14, 16, and total sample (1.6 + Xc + 1 + 36); age 7 (1.7 + Xc + 1 + 32); and age 15 (1.5Xc + 1 + 40).

Seven-Subtest Short Forms

One strategy is to delete the most time-consuming subtests and give as many of the shorter subtests as possible. J. Ryan and Ward (1999) developed a WAIS-III seven-subtest short form (Information, Digit Span, Arithmetic, Similarities, Picture Completion, Block Design, Digit Symbol-Coding), which takes 35 minutes to administer. A slight variation from this short form is to substitute Matrix Reasoning for Block Design. This has the advantage of providing a slightly more accurate estimate of Performance IQ

(Axelrod, Ryan, & Ward, 2001). The subtest scores can be prorated and the resulting scores can be used to develop estimates of Full Scale, Verbal, and Performance IQs. Alternatively, tables provided in Sattler (2001, p. 835) and J. Ryan (1999) can be used to develop estimated IQ scores. Performance and Full Scale IQ scores have been found to be nearly as reliable as for full-administration IQs with the average Full Scale standard error of measurement being 2.80 (and 2.72 for the version with Matrix Reasoning) versus 2.58 for the full WAIS-III Full Scale IQ (J. Ryan & Ward, 1999). Correlations between the J. Ryan and Ward (1999) seven-subtest short form and a full administration were .98 for the Full Scale IQ, .97 for the Verbal IQ, and .95 for the Performance IQ (.96 using Matrix Reasoning). Thus, the psychometric properties of the seven-subtest short form are excellent, and administration times are only marginally longer than for the Vocabulary-Arithmetic-Block Design-Picture Arrangement four-subtest short form.

The Satz-Mogel/Yudin Short Forms

An alternative to administering various combinations of subtests is to use every subtest but limit the number of items from each of the subtests. The most frequently used variation is the Satz and Mogel (1962) approach, which was originally developed for the WAIS but can also be used for the WAIS-III and WISC-III. The procedure is to administer every third item for Information and Vocabulary and multiply the scores by three to obtain the raw scores. Only odd items are administered for Similarities, Arithmetic, Comprehension, Block Design, Object Assembly, and Picture Completion, and each score is multiplied by two to obtain the respective scaled scores. Full administrations are given for Digit Span, Digit Symbol-Coding, Letter-Number Sequencing, Matrix Reasoning, and Symbol Search. The entire procedure takes approximately 40 minutes and the derived IQs have correlations similar to the best four-subtest variations. A distinct advantage over four-subtest variations is that the Satz-Mogel approach samples a wider range of areas. This is likely to increase the stability of scores over a wider variety of populations and allows clinicians to develop inferences over a larger number of behaviors. Research with the WAIS-III has indicated that IQs derived from the Satz-Mogel usually did not vary more than 6 points when compared with the full administration (J. Ryan, Lopez, & Werth, 1999). In addition, a full 86% of the clients had the same IQ classifications. A caution is that, although a score is provided for each subtest, it is inappropriate to attempt a profile analysis because the individual subtests are not sufficiently reliable (J. Ryan, Lopez, & Werth, 1999).

A WISC-R/WISC equivalent of the Satz-Mogel approach was developed by Yudin (1966) and has the same advantages and disadvantages and follows a nearly identical procedure. If adapted for the WISC-III, Digit Span, Mazes, and Symbol Search would not be administered because they are optional subtests; but Coding, like Digit Symbol-Coding on the WAIS-III, would be given in its entirety. However, if examiners did decide to use the Symbol Search subtest because of its good psychometric properties or clinical relevance, it would need to be given in its entirety.

Modified Format

A final approach is the elimination of early, easy items on each of the subtests. This is most appropriate for relatively bright subjects but should be used cautiously with

persons of below-average intelligence. Cella (1984) has provided WAIS-R guidelines for the number of items to be omitted based on a subject's performance on the Information subtest. Research on the WAIS-III using this format has not yet been conducted. However, such a procedure with the WAIS-R has been found to have an almost exact correlation (.99) with a full administration and yet can reduce the total administration time by 25%. Despite this high correlation, some caution should be exercised toward Cella's Modified Format and the Satz-Mogel approaches. First, lowered internal consistency is likely to reduce subtest reliabilities sufficiently to render profile analysis questionable. Second, examinees are disadvantaged because they are not able to have as many previous subtest items to practice on (as items are skipped) before being administered more difficult items. The result may be that the norms for the full administration may not necessarily apply to the shortened versions.

RECOMMENDED READING

Gregory, R. J. (1999). *Foundations of intellectual assessment: The WAIS-III and other tests in clinical practice.* Boston: Allyn & Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2002). *Assessing adolescent and adult intelligence* (2nd ed.). Boston: Allyn & Bacon.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.

Chapter 6

WECHSLER MEMORY SCALES

The Wechsler memory scales are one individually administered, composite batteries designed to better understand various components of a patient's memory. Now in its third edition (WMS-III), it has been co-normed with the WAIS-III. Another major feature is that it provides a full range of memory functioning and has been carefully designed according to current theories of memory. As a result of these features, it is typically considered to be a core component of any thorough cognitive assessment, which is reflected in its being ranked as the ninth most frequently used test by clinical psychologists (and third by neuropsychologists; Camara et al., 2000).

Memory complaints are extremely prevalent among client populations. They are associated with depression, anxiety, schizophrenia, head injuries, stroke, learning disabilities, and neurotoxic exposure. For example, the impact of alcohol and other drugs on memory might need to be carefully evaluated. Occupational contexts might need to similarly evaluate the impact of workers who have been exposed to industrial agents (lead, mercury, organic solvents) that can potentially result in impaired memory function. The increasingly aging population means that distinguishing normal memory loss from the early expression of dementia will become progressively more important. One crucial differential diagnosis is to distinguish between pseudodementia resulting from depression and Alzheimer's disease. As various drugs are developed for treating cognitive difficulties, it will also become increasingly important to monitor client improvement with a particular emphasis on memory functions. This array of symptoms suggests a developmental perspective in that children are most likely to experience memory complaints related to learning disabilities, adults typically experience difficulties because of neurotoxic exposure or head injuries, and older populations have memory problems related to dementing conditions.

Many of the early conceptualizations of memory considered it a unitary process. From a practical assessment perspective, it was not necessary to have a composite battery that assessed various components of memory. In contrast, more recent conceptualizations consider memory to have various components. One major distinction is between short-term and long-term memory (sometimes described as primary and secondary memory storage, respectively). For memory to be effectively stored, there also needs to be some active engagement on the part of the person. Thus, "working memory" was conceptualized as containing an executive component that initiated, monitored, and evaluated information. It also included an attentional component that had a limited capacity. A further well-supported distinction is between memory that is conscious and reflected in verbal reports of facts, events, and experiences (*declarative, explicit,* or *episodic* memory) versus memory that is more unconscious and measured implicitly by changes in

performance (*procedural,* or *implicit* memory). Finally, memory can involve various sensory components, particularly visual and auditory modes of processing.

HISTORY AND DEVELOPMENT

In some ways, the development of the Wechsler memory scales have paralleled the development of knowledge on memory. Each of the three editions has increasingly incorporated advances in the theoretical understanding of memory. The original Wechsler Memory Scale (WMS; Wechsler, 1945) reflected the earlier nonspecific conceptualizations of memory. It was composed of brief procedures on memory for number sequences, text, simple visual designs, and paired words. The advantage of using a variety of procedures was that a client might have intact memory for visual information but not auditory information or vice versa. Despite the fact that the early WMS procedures could be *logically* divided into visuospatial versus auditory tasks, the overall scoring was a composite Memory Quotient that, similar to the Wechsler intelligence scale IQs, had a mean of 100 and a standard deviation of 15. This was extremely valuable information for clinicians because they could easily compare a client's Intelligence Quotient with their Memory Quotient. Any large discrepancy could be investigated further to understand the underlying reasons for such a discrepancy. The WMS was also quite popular as it was a relatively brief procedure, typically taking about 15 minutes to complete. Because retesting a client would be likely to result in practice effects, it had the further advantage of having a parallel form. As a result of these advantages, it became an ubiquitous procedure among clinicians.

The WMS had surprising longevity given a formal new version did not become available until 1987 (a 42-year interval). The WMS was limited, however, because it included unsophisticated methods of scoring the various procedures. In addition, the algorithms to determine the Memory Quotient were overly simple because they did not consider a sufficient number of client variables. The norms were derived from a small sample of 200 patients between ages 25 and 50 at Bellevue Hospital. Scores for either older or younger persons were extrapolated from this sample but were not based on actual participants. In addition, the alternate form was rarely used, and the research supporting it was quite limited. Finally, it did not reflect advances in knowledge related to memory processes.

One early attempt to correct for the deficiencies of the WMS was Russell's (1975, 1988) adaptation in which he administered two of the subtests (Logical Memory and Visual Reproduction) in an immediate format combined with a delay of 30 minutes. This allowed comparisons to be made between short-term and long-term memory. Research on Russell's WMS supported the predicted difference between left (relatively lowered auditory recall based on Logical Memory) and right (relatively lowered visual reproduction based on Visual Reproduction) hemisphere lesions. Despite these advantages, the psychometrics were weak and it was poorly standardized. Unfortunately, it was titled *Wechsler Memory Scale-Revised* (WMS-R). This potentially could create confusion because The Psychological Corporation developed a full revision of the WMS that was also titled *Wechsler Memory Scale-Revised.* Subsequent publications have attempted to clarify the two versions by referring to them as either *Russell's WMS-R* or the *WMS-R.*

The 1987 revision (Wechsler Memory Scales-Revised or WMS-R) was a significant improvement over the WMS. It had age-related norms for nine different age groups ranging between 16 and 17 years for the youngest group and 70 to 74 years for the oldest group. However, three of the age groups (18 to 19 years, 25 to 34 years, and 45 to 54 years) were not based on actual subjects but rather were estimated based on the gradual monotonic decrease in performance. The standardization sample was composed of 316 persons, who had characteristics that closely approximated 1980 census data. There were approximately 50 subjects in each of the age groups. Whereas the WMS had only one composite Memory Quotient, the WMS-R had twelve subtests from which the following five composite scores could be derived: General Memory, Attention-Concentration, Verbal Memory, Visual Memory, and Delayed Recall. Each of the index scores has a mean of 100 and a standard deviation of 15. This division into index scores is consistent with theories that have divided memory into short term and long term (note the Delayed Recall used to assess long-term memory) and verbal/auditory versus visual (note the Verbal Memory and Visual Memory indexes).

Reliability of the WMS-R has been generally low to adequate. Test-retest reliabilities over a four- to six-week interval were moderate (i.e., Mental Control $r = .51$; Visual Paired Associates $r = .58$; Verbal Paired Associates $r = .60$ for initial and .41 for delayed recall). Internal consistencies ranged from a low of .44 for Figural Memory to a high of .88 for Digit Span (Wechsler, 1987). The standard error of measure ranged from a high of 8.47 for the Visual Memory Index to a low of 4.86 for the Attention-Concentration Index (Wechsler, 1987).

Similar to studies on reliability, the validity of the WMS-R has been good to adequate. A number of factor analytic studies have generally found that the different subtests can be organized into two major factors described as a General Memory/Learning factor and an Attention-Concentration factor (Bornstein & Chelune, 1988; Roid, Prifitera, & Ledbetter, 1988; Wechsler, 1987). There was also some evidence for a three-factor solution composed of Verbal Memory, Nonverbal Memory, and Attention (Bornstein & Chelune, 1988). A wide range of studies supports the ability of the WMS-R to distinguish between normal and clinical groups (A. Hawkins, Sullivan, & Choi, 1997; Reid & Kelly, 1993; Wechsler, 1987), distinguishes the relative severity of deficits based on subjective complaints (Gass & Apple, 1997), provides an index that relates to client ratings of level of everyday memory (Reid & Kelly, 1993), and predicts the degree of brain atrophy (Gale, Johnson, Bigler, & Blatter, 1995). In addition, the Attention-Concentration Index was found to be one of the most sensitive measures in identifying cognitive impairment (M. Schmidt, Trueblood, Merwin, & Durham, 1994). Despite a conceptual basis for believing that visual and verbal memory would relate to laterality of deficits, research on this has produced inconsistent results (Chelune & Bornstein, 1988; Loring, Lee, Martin, & Meador, 1989). Therefore, interpretations related to laterality should be made with caution. For example, having an impaired Visual Memory Index but good Verbal Memory Index does not necessarily mean that a patient has unilateral damage to the right hemisphere.

The WMS-R had clear advantages over the WMS because it had a far better normative base, was validated on diverse populations, had quite extensive studies performed on it, and divided memory into various indexes, thereby allowing the possibility for measuring various aspects of memory. It was, however, its weaknesses

that resulted in its revision within a relatively short period. One of the most serious limitations of the WMS-R has been the relatively low reliabilities of the subtests and indexes (Elwood, 1991). This is likely to significantly reduce the accuracy of measurements. In addition, the different indexes are probably not very good measures of specific components of memory. This is not to say they are not sensitive to both *general* cognitive impairment and the *degree* of that impairment. However, the specific nature of the impairment cannot be accurately determined by referring to the specific indexes despite the fact that the names of the indexes suggest that this differentiation can be made. Finally, current theories of memory were not used in the design of the WMS-R (Lichtenberger, Kaufman, & Lai, 2002).

The Wechsler Memory Scale-III was published just ten years after the release of the WMS-R. The new revision was designed not merely as a facelift of the WMS-R, but rather a "state of the art assessment instrument that comprehensively addresses the complexity of brain/behavior relationships involved in learning and memory" (Edith Kaplan in the forward to the WMS-III manual, p. iii). To accomplish this goal, new subtests were added, scoring procedures were made more sophisticated, stimulus materials were changed, and new index configurations were developed. This resulted in six primary and five optional subtests. Eight index scores could then be developed (see Table 6.1). Whereas the manual states that it is possible to administer the six primary

Table 6.1 WMS-III indexes, primary subtests, and optional subtests

Indexes	Subtests used to Calculate Indexes
Auditory Immediate	Logical Memory I, Verbal Paired Associates I
Visual Immediate	Faces I, Family Pictures I
Immediate Memory	Logical Memory I, Verbal Paired Associates I, Faces I, Family Pictures I
Auditory Delayed	Logical Memory II, Verbal Paired Associates II
Visual Delayed	Faces II, Family Pictures II
Auditory Recognition	Logical Memory Recognition, Verbal Paired Associates Recognition
General Memory	Logical Memory II, Verbal Paired Associates II, Faces II, Family Pictures II, Auditory Recognition
Working Memory	Letter-Number Sequencing, Spatial Span
Optional Subtests and Procedures:	Information and Orientation
	Word Lists I and II
	Visual Reproduction I and II
	Mental Control
	Digit Span

Source: Adapted from "The Wechsler Memory Scales," by Franzen and Iverson, 2000. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration.* New York: Wiley.

subtests in 30 to 35 minutes, research with a clinical population indicated that it took 42 minutes to administer the eleven primary subtests (see Axelrod, 2001).

One of the most important aspects of the WMS-III is that it was developed simultaneously with the WAIS-III. This has enabled the two tests to not only share two subtests, but also to be co-normed. The normative sample consisted of 1,250 adults ranging between 16 and 89 years. Instead of 9 groups as in the WMS-R, the WMS-III had 13 different groups. These groups not only had more subjects (50 in each group for the WMS-R versus 100 for the first 11 groups of the WMS-III), but also extended to a far higher age range (74 for the WMS-R versus 89 for the WMS-III). This is appropriate because one of the more important functions of memory assessment is to evaluate older clients.

RELIABILITY AND VALIDITY

The WMS-III has better reliability than its predecessor, the WMS-R. The *WAIS-III/WMS-III Technical Manual* indicates that internal consistency for the primary subtest scores ranges between .74 and .93 for all age groups. As would be expected, the primary indexes have even better internal consistencies of .82 or higher. The one exception is the somewhat lower reliability of .74 for Auditory Recognition Delayed. Test-retest reliabilities for all age groups over a two- to twelve-week interval mostly ranged between .62 and .82 for the individual subtests and between .75 and .88 for the indexes. Again, Auditory Recognition Delayed had a somewhat lower reliability of .70. The technical manual states that even those subtests requiring the most judgment (Logical Memory I and II, Family Pictures I and II, Visual Reproduction I and II) had interscorer reliabilities above .90. However, scorers (on the WMS-R) have been found to make an average of four errors per protocol, indicating that extra care should be taken to ensure that scores are congruent with the criteria in the manual (Sullivan, 2000).

There is ample evidence that the WMS-III can effectively differentiate between clinical and normal populations. Various clinical groups (Alzheimer's disease, Huntington's disease, Parkinson's disease, multiple sclerosis, chronic alcohol abuse, temporal lobe epilepsy, schizophrenia) consistently score lower than the standardization sample (D. Fisher et al., 2000; K. A. Hawkins, 1998; The Psychological Corporation, 1997). For example, patients with mild Alzheimer's disease scored in the 60 to 71 range for most of the primary indexes except for a mean score of 80 for Working Memory (Psychological Corporation, 1997). Similarly, Fisher et al. found that patients with moderate-to-severe traumatic brain injury scored low on all indexes. WMS-III Visual Delayed and Visual Immediate (and WAIS-III Processing Speed) were particularly sensitive to the severity of the injury. Finally, the WMS-III has been found to correspond to clinician ratings of the severity of brain injury (Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999).

Although differentiating between normal and clinical groups is essential, it is also a relatively easy criterion to achieve. What is particularly crucial for the practicing clinician is to determine whether the individual indexes can accurately measure subcomponents of memory. Factor analytic studies and determining whether patterns of scores

match theories of memory (i.e., visual/verbal in relation to laterality) are particularly important. The technical manual reported a factor analysis using the standardization sample and concluded that, for ages between 16 and 29, a three-factor model composed of working memory, visual memory, and auditory memory most closely fit the data. In contrast, a five-factor model composed of working memory, auditory immediate memory, visual immediate memory, auditory delayed memory, and visual delayed memory fit the age groups from 30 to 64 and 65 to 89. For ages 30 to 89, this closely corresponds to five of the eight index scores. The change in factor structure between the younger and older age groups is also consistent with findings that the components of memory become more clearly distinguishable ("dissociated") with age (Dolman, Roy, Dimeck, & Hall, 2000). Thus, the index scores might become more meaningful with older populations. An additional factor analysis also using the standardization sample supported the three-factor model reported in the technical manual composed of working memory, visual memory, and auditory memory (Millis, Malina, Bowers, & Ricker, 1999). Although the researchers did not find support for the five-factor model, it should be noted that they did not separate their groups into different ages. It is interesting to note that the previous WMS-R consistently produced factors that supported a distinction between immediate memory and delayed recall (Bowden et al., 1997; Hunkin et al., 2000). This suggests that the WMS-III may have succeeded in emphasizing more visual tasks (as well as more clearly defined factors) but, when compared with the WMS-R, may have lost differentiation between immediate and delayed memory (K. Hawkins, 1998; Millis et al., 1999). In addition, the Immediate Memory and General Memory indexes may be redundant because they have been found to correlate .98 in a variety of clinical populations (K. Hawkins, 1998; Weiss & Price, 2002).

There is some evidence that various types of clinical populations perform differently on various indexes. As would be predicted given knowledge about brain laterality, patients with right lobectomies performed considerably lower on the Visual Immediate Index than on the Auditory Immediate Index. Conversely, left lobectomy patients performed worse on the Auditory Immediate Index when compared with their Visual Immediate Indexes (K. Hawkins, 1998). However, both groups of patients performed poorly on the Visual Immediate Index. Both the Visual Immediate and Visual Delayed Indexes have also been found to be relatively sensitive to alcohol abuse, multiple sclerosis, and traumatic brain injury (D. Fisher et al., 2000; K. Hawkins, 1998). Thus, visual memory may be a particularly sensitive indicator of brain injury in general. Somewhat similarly, traumatic brain-injured patients with mild injuries showed lower than expected scores on Auditory Immediate, Visual Immediate, Visual Delayed, and General Memory (D. Fisher et al., 2000). With more severe injury, Visual Delayed and Visual Immediate were particularly affected. The index least susceptible to lowering was the Auditory Recognition Delayed.

The technical manual reports a number of performances for clinical populations. For example, Korsakoff's syndrome is characterized by severe difficulties with encoding and storing new information but the patient's attention and working memory are normal. This is reflected on the WMS-III index performances wherein Working Memory was in the normal range but all other index scores were in the impaired range (Psychological Corporation, 1997). The previous sampling of studies indicates that many of the predicted theoretical and clinical patterns of performance have occurred on the various WMS-III indexes.

ASSETS AND LIMITATIONS

The WMS-III is generally an excellent instrument capable of measuring a wide range of memory functioning. It has been based on theoretical research into the processes of memory, it has excellent standardization, and most research indicates solid empirical support. It is clearly an improvement over both the original WMS and the WMS-R. Perhaps its strongest advantage is its co-norming with the WAIS-III. This allows practitioners to make realistic comparisons between performance on the two instruments.

An important unanswered question with the WMS-III is the extent it can actually measure the various components of memory. Its divisions (and corresponding indexes) into working, visual, and auditory memories are quite well supported. However, the distinction between immediate and delayed memory may be questionable. In addition, the Immediate and General Memory indexes may be redundant. Thus, the number and titles of the indexes may promise more specificity than can actually be delivered. A related and important issue is that the various components of memory (and corresponding indexes) are likely to perform differently across various clinical populations and age groups. A final unanswered question in need of further exploration is the extent to which the WMS-III relates to aspects of everyday memory. Given the considerable research that resulted from the WMS-R, these, and many additional questions, will be answered over the next few years.

The original WMS had the advantage of taking only 15 minutes to administer. The WMS-R and now the WMS-III have increased the administration times to an average of 42 minutes, but it may actually take up to 100 minutes for some clinical populations (Lichtenberger et al., 2002). When the WMS-R was released, many clinicians either continued to use the WMS, or used only selected portions of the WMS-R (Butler, Retzlaff, & Vanderploeg, 1991; Sullivan & Bowden, 1997). Given the present co-norming with the WAIS-III and vastly superior standardization, it is difficult to justify the continued use of the original WMS. It is, however, quite likely that practitioners concerned with time efficiency will use only portions of the WMS-III. For example, they might give only those subtests that are necessary to develop an Immediate Memory index and then compare this with the WAIS-III IQs to notice discrepancies. A further option might be to give only those subtests that seem to be most sensitive to cognitive impairment (Visual Immediate and Visual Delayed) or to use empirically based short forms to extrapolate various index scores. For example, a three-subtest short form consisting of Logical Memory, Verbal Paired Associates, and either Faces or Family Pictures correlated at a .97 level with General Memory (and Immediate Memory; Axelrod, Ryan, & Woodward, 2001). A two-subtest short form composed of Logical Memory and Verbal Paired Associates had a quite similar correlation of .96 with General Memory (and Immediate Memory). These two short forms account for 95% to 97% and 87% of the variance in General Memory and Immediate Memory, respectively (Axelrod & Woodward, 2000). However, this use of nonstandardized administrations raises the possibility of introducing unknown error.

The scoring and administration of the WMS-III is mostly clearly described in the manual. However, Logical Memory does not present guidelines regarding the speed at which the stories should be read. It also does not have guidelines for intonations, pauses, or inflections. Examiner variation in each of these areas may, therefore, result in the potential for error. Lichtenberger et al. (2002) suggested that an audiotaped

administration might be introduced by the test developers. A further issue with both Logical Memory I and II is its high degree of cultural loading; therefore, persons whose first language is not English may be disadvantaged.

In many ways, the complexity of the WMS-III is an advantage because it allows for the possibility of assessing a wide range of memory functions. It should be noted that other comprehensive batteries have similar levels of complexity. However, the fairly complex procedures may discourage some clinicians from learning and using it. The relatively long administration and detailed scoring procedures may also introduce the possibility of scoring and computational errors.

INTERPRETATION PROCEDURE

The WMS-III measures a wide range of different functions. As a result, interpretation can be complex. The following sequence is a general strategy that is conceptually quite similar to the interpretive procedure for the WAIS-III. The initial interpretations are for global measures, followed by progressively more specific considerations. In addition, far more interpretive strategies might be used. For example, scoring for four Auditory Process Composites are provided in the *WMS-III Administration and Scoring Manual.* These indicate a patient's level of initial learning (Single-Trial Learning), rate which learning improves over multiple trials (Learning Slope), degree information is retained after a delay (Retention columns), and extent cueing increases a person's retrieval of information (Retrieval Composite). In addition, interpretation of subtests has not been included because neither the psychometrics nor the research literature warrants such procedures. Far more in-depth strategies can be found in Lichtenberger et al. (2002). The information included next is considered both manageable and psychometrically sound and thus provides a solid, clear introduction to WMS-III interpretation.

1. Interpret the General Memory Index

The original Wechsler Memory Scale resulted in a single memory quotient that clinicians found quite useful because they could compare it to a person's Full Scale IQ to note whether there were any large discrepancies. This information could then be used to infer a relative strength or weakness in global memory compared to a person's other cognitive abilities. The General Memory Index can be used in the same manner. Indeed, an IQ score that is much larger than a global memory score (IQ > Memory) has often been used to "red flag" the possible presence of brain dysfunction. This is based in part on the clinical observation that one of the most frequent complaints by patients with brain damage is that they have memory difficulties.

One item that may not be immediately clear is the title "General Memory," which does not clearly indicate that it measures delayed memory. However, a brief review of the subtests comprising the General Memory Index indicates that they are all delayed tasks. Part of the rationale for referring to it as "general" memory is that the types of tasks assessed by this index (delayed tasks) relate more clearly to everyday types of memory functions. As a result, a more accurate title might have been the "Global Delayed Memory" index (Lichtenberger et al., 2002). Conceptually, this can still be

considered an adequate global measure of memory given that memory refers to the *persistence* (delay) of learning, whereas more immediate measures of "memory" are greatly influenced by attention. Thus, because each of the subtests used to calculate the General Memory Index involves delay/persistence, any lowering of performance because of poor attention will result in a corresponding lowering in performance on the General Memory Index.

Despite the potential usefulness of the IQ > Memory distinction, caution should be used for two major reasons. First, large fluctuations occur among normal populations. Differences of 13 points occurred in 15% of the standardization sample and 15 points in 10% of the sample. An abnormal discrepancy (occurring in only 5% of the standardization sample) was a difference of 20 points or more. Table 6.2 summarizes this information and also describes the extent of unusual differences for the other index scores. Second, an IQ > Memory difference has been found as only a weak indicator of dysfunction. This is because, with brain dysfunction, there is often a corresponding decline in other nonmemory abilities as well. This results in a reduction in not only IQ, but also in measures of memory (General Memory Index), which thereby results in little difference between the two measures.

Because quite large differences between IQ and Memory are a fairly common occurrence even among normal populations, a more sensitive indicator might be to compare the relatively stable WAIS-III Verbal Comprehension Index with the highly sensitive WMS-III Visual Immediate Memory Index (K. Hawkins, 1998). While the frequency of large differences is not provided in the manual, differences of 18 or more should be investigated further. It also suggests that further research may find this a particularly good indicator of brain dysfunction.

2. Interpret Patterns of High and Low Index Scores

The purpose of interpreting patterns of index scores is to better understand a person's relative memory-related strengths and weaknesses. Initially, this might be done by noting the absolute values of the index scores. For example, a relatively low score on Visual Memory might indicate a relative weakness in this modality. In contrast, a low

Table 6.2 Reliable, unusual, and abnormal differences between FSIQ and Index Scores averaged for all ages

		Unusual		Abnormal
	Reliable	≤15%	≤10%	≤5%
Auditory Immediate	8.8	14	17	23
Visual Immediate	13.1	18	22	29
Immediate Memory	9.7	15	17	23
Auditory Delayed	11.7	14	17	23
Visual Delayed	12.9	17	21	26
Auditory Recognition Delayed	15.7	16	20	25
General Memory	9.9	13	16	22
Working Memory	11.9	13	15	20

Derived from Table C.1 and C.4; pp. 288 and 291 in the Technical Manual.

score on Working Memory might suggest attentional difficulties and possibly problems with organizing information in sequential order. However, the same caveats for interpreting index/subtest fluctuations that were discussed for the Wechsler intelligence scales are also relevant for the WMS-III. Specifically, base rates for discrepancies need to be considered so that relatively frequently occurring differences are not overinterpreted to indicate pathology. In addition, the indexes may lack sufficient specificity. Clinicians should also be aware that fluctuations could occur for a number of different reasons. It is up to each clinician to carefully evaluate these various possibilities by carefully integrating additional relevant information. Therefore, the following possible interpretations should be considered tentative.

The level of significance between the various patterns of indexes should be determined first. This can be accomplished by subtracting one index scale score from another and consulting Table F.1 in the *WMS-III Administration and Scoring Manual.* For example, a difference of 14.5 points between Auditory Immediate and Visual Immediate was significant at the .05 level (across all age ranges). Table 1.2 (also in the administration and scoring manual), however, indicates that although this is statistically significant, a 14.5-point difference still occurred among 25% of the standardization sample. When the level of significance and frequency of this difference is established, clinicians can consider possible interpretations.

The following clusters of profile interpretations are organized according to those distinctions that are both most clinically useful and have received empirical and theoretical support. One of the major concerns is to know whether there are differences between immediate (short-term) and long-term (delayed) memory. A further pattern that sometimes emerges is the difference in the relative strength of visual or auditory modalities. It is also often relevant to know if there are differences between a person's ability to retrieve (recall) information or the less difficult task of recognizing this material when presented with relevant stimuli. The final distinction is between complex attentional processes (working memory) that involve manipulating newly presented information as opposed to simpler encoding and acquisition. Knowledge related to each of these components of memory has relevance for diagnosis, treatment planning, as well as for understanding normal levels of strengths and weaknesses.

Immediate/Delayed

Immediate Memory/General Memory (Delayed) As noted previously, the General Memory index is most appropriately considered a measure of general delayed memory (or Global Delayed Memory; Lichtenberger et al., 2002). Thus, it can be used as the comparison score to contrast a person's immediate memory with his or her delayed memory. This is an important distinction that concerns practicing clinicians. As a result, it may even be one of the referral questions. If delayed memory is considerably lower than immediate memory (12 points or more for a .05 level of significance), it suggests that the person can initially learn material but then the information decays over a period of time. It should be stressed in this regard that performance on immediate memory becomes the benchmark for how much information has been lost. In other words, unless a person has at least learned something initially, there is nothing to lose. The exception might be that a person has acquired information but then may not be able to recall it (poor retrieval). However, recognizing the information is generally a much easier

task. This means that the person might be able to accurately recognize information he or she may not have otherwise been able to recall/retrieve without the prompting (see interpretations related to Auditory Delayed/Auditory Recognition Delayed).

A further issue is that factor analysis of the immediate/delayed distinction on the WMS-III may not be as strong as would be optimal (see Psychological Corporation, 1997 versus K. Hawkins, 1998; Millis et al., 1999). It is likely that the two indexes (Immediate Memory and General Memory) are redundant, as they had a .98 correlation for a variety of clinical populations (K. Hawkins, 1998; Weiss & Price, 2002). Thus, WMS-III interpretations related to immediate versus delayed memory should be tentative.

Auditory Immediate/Auditory Delayed This discrepancy occurs in the same modality (auditory) but evaluates whether there are differences between short-term and long-term abilities. Low scores on both of these indexes relative to a person's IQ suggest difficulties in verbal learning and memory. In contrast, if a person's Auditory Immediate index is significantly higher than his or her Auditory Delayed (13 points for a .05 level), he or she might be experiencing a high rate of forgetting. For example, the person might be told about a meeting time and place or given a set of instructions but would have difficulties retaining this information. However, this inference needs to always take into account how much he or she originally learned based on the height of Auditory Memory as this is an indication of how much information was originally acquired. In other words, the person can forget information only in relation to how much was originally learned.

Visual Immediate/Visual Delayed Visual Immediate and Visual Delayed are within the same modality (visual), but the difference is short-term versus longer term differences in ability within this modality. Low scores on both indexes relative to a person's intelligence would indicate an overall memory difficulty with this modality. However, if immediate memory is significantly higher (17 points for a .05 level), there is likely to be important losses of visual information over a period of time. For example, the person might have studied a map or been to a meeting; but, after a period of time, he or she may not be able to recall relevant directions or remember who had attended the meeting. Keep in mind that the Visual Immediate score is always the benchmark for comparing Visual Delay because the Visual Immediate index is dependent on how much the person originally learned.

Modalities (Auditory/Visual)

Auditory Immediate/Visual Immediate One of the basic distinctions supported by WMS-III factor analysis is between auditory and visual memory. The difference between these modalities (and the indexes that measure them) can thus be used to hypothesize relative auditory versus visual strengths and weaknesses. Thus, a significant difference between Auditory Immediate and Visual Immediate (11 points or more for a .05 level of significance) can indicate either lifelong patterns related to differences in abilities or acquired deficits in these modalities. Laterality differences have been previously noted (K. Hawkins, 1998) in that patients with unilateral left hemisphere damage have been found to do more poorly for verbal-auditory information than for visual

information. For example, they would be expected to have particular difficulty when given verbal directions. In contrast, they might perform far better when shown a visual map of how to get from one place to the next. In contrast, patients with unilateral right hemisphere damage would be expected to do more poorly on visual immediate tasks. Thus, they would be expected to benefit most from auditory-verbal directions compared to directions that were visually presented. However, visual memory performance was found to be the most sensitive to any type of brain damage, and patients with both unilateral right and left hemisphere damage performed poorly on visual memory types of tasks. If one modality was found to be relatively stronger than another, this stronger modality might be used to maximize learning. For example, if a person's auditory learning was poor, he or she might use learning strategies that capitalized on visual modes (or vice versa).

Auditory Delayed/Visual Delayed The same interpretive considerations as noted previously (between Auditory Immediate/Visual Immediate) also apply, except that the extent that memory is lost over a period of time is also measured. Thus, a significant difference between the two indexes (12 or more points) may indicate that there is more "decay" of memory in one modality than in another. Again, this may have practical implications in terms of developing learning strategies because either the auditory or visual mode might be used to compensate for a relative weakness.

Recall (Retrieval) versus Recognition

Auditory Delayed/Auditory Recognition Delayed Distinguishing whether a person once knew something or never knew it can be important clinical information. For example, patients with early dementing conditions frequently complain of difficulty retrieving information. Relevant behavioral observations might be statements such as "I know the answer but I just can't think of it." Qualitative approaches to the WAIS-R and WISC-III use multiple-choice formats on the Wechsler intelligence scale Information or Vocabulary subtests to try to determine this (see E. Kaplan et al., 1999; Milberg et al., 1996). The WMS-III uses this strategy by presenting recognition items on recall of stories (Logical Memory II) and pairs of words (Paired Associate Learning II). Significantly higher scores on recognition compared with delay (16 points or more for a .05 significance) suggest that the client has retrieval difficulties. He or she might experience this as frustration over not being able to find the correct word or difficulty recalling relevant and commonly known facts. Friends or work colleagues might have commented that he or she seems to have difficulty remembering information that other people are certain he or she once knew.

Complex Attention versus Acquisition/Encoding

Working Memory/Immediate Memory The WMS-III Working Memory Index is similar to the WAIS-III index of the same name in that they share Letter-Number Sequencing. However, the WAIS-III is composed of two auditory tasks (Digit Span and Letter-Number Sequencing), whereas the WMS-III has one auditory task (Letter-Number Sequencing) and one visually presented task (Spatial Span). This similarity is reflected in that they are highly correlated (.82), and they measure similar functions (see Working Memory interpretations in Chapter 5). Specifically, Working Memory

measures a person's ability to attend to stimuli while at the same time performing other mental tasks. It also requires the person to effectively sequence information and make mental shifts. Because a person must monitor his or her performance and respond flexibility, Working Memory is also related to executive functioning. Thus, it involves complex attentional abilities. In contrast, the tasks of Immediate Memory (recalling story events, recognizing details from pictures) do not require as high a level of attention.

If Working Memory is significantly lower than Immediate Memory (14 points for the .05 level of significance), it suggests that the person can adequately work with simple encoding and acquisition, but may have more difficulty with more complex attentional tasks. For example, a brain-injured person might be able to learn relatively simple information in a quiet environment, but when confronted with distractions ("multitasking"), he or she might be unable to cope effectively. If both Working Memory and Immediate Memory are low in relation to intelligence, it suggests that poor attention is affecting the person's ability to learn new information.

Working Memory/General Memory A Working Memory Index that is significantly lower (10 points or more for a .05 level of significance) than the General Memory Index indicates the person is likely to experience difficulties with tasks requiring complex attention. In contrast, he or she is likely to more effectively work with tasks requiring simple encoding and acquisition. Thus, some of the same interpretations that apply for significant differences between Working Memory/Immediate Memory also apply for Working Memory versus General Memory because General Memory comprises the same tasks as Immediate Memory, but there is a 25- to 35-minute delay. In contrast, a General Memory that is significantly lower than Working Memory suggests that complex attentional skills are a relative strength. It may also suggest that there has been some decay of memory between the immediate tasks and delayed tasks (check the difference between Immediate Memory and General Memory).

3. Evaluate More Specific Functions Derived from Individual Subtests

Research and clinical lore on interpreting the WMS-III subtests are minimal when compared with the Wechsler intelligence scales. The advantage for WMS-III interpreters is that it helps professional psychologists stay more clearly focused on the much better validated index scores. However, the various combinations of subtests do measure a number of functions that are not necessarily extracted by the index scores. These might include visual construction, degree to which the patient is oriented, and visual perception. A listing of some of these functions, along with additional interpretive material and relevant WMS-III subtests, follows (derived from Franzen & Iverson, 2000 and Lichtenberger et al., 2002):

- *Orientation:* The degree to which the person is oriented to person, place, date, and general information. This is core information for any mental status evaluation (see optional Information and Orientation section).

- *Simple Attention:* Ability to comprehend and repeat simple auditory and visual information. This is assessed on the WMS-III by procedures that request the patient to repeat a series of numbers and repeat a series of sequential taps or a form board (see Digit Span Forward and Spatial Span Forward).
- *Complex Attention:* Ability to concentrate on, hold, organize, and manipulate complex information. WMS-III-related tasks include repeating digits backward, repeating a sequence of taps on a form board backwards, mentally reorganizing a series of numbers mixed in with letters (see Digit Span Backward, Spatial Span Backward, and Letter-Number Sequencing).
- *Learning:* Ability to combine new information and later recall it. Specific tasks include learning pairs of words that belong together and recalling a list of words that has been read by the examiner (see Verbal Paired Associates and Word Lists).
- *Visual Perception:* Ability to visually distinguish between and match geometric designs (see Visual Reproduction Discrimination).
- *Visual Construction:* Ability to accurately perceive and copy designs (see Visual Reproduction Copy).
- *Malingering:* The Logical Memory Delayed Recognition task requests clients to state whether (yes or no) an item was included in one of the previously read stories. Because random guessing would produce a score of 50%, scores of less than this suggest that the client is malingering (see Killgore & Dellapietra, 2000). Malingering may also be suggested if recognition doesn't improve in comparison to recall (e.g., negative Retrieval Total Score) because recognition tasks are easier than free recall tasks. Malingering may also be indicated if a patient has a lower score on the Logical Memory I Thematic Score than on the Logical Memory I Total Score. The patient should be able to have a higher performance on the far easier Thematic Score (recalling the underlying themes of stories) than the more difficult Total score (that requires him or her to recall quite specific units of information). A final quite general indicator is dramatic differences between a person's day-to-day functioning (based on evidence from corroborating sources) and performance on WMS-III measures.

In addition to the previous listing of nonindex functions, a number of observations have been made regarding behaviors and performance on the individual subtests. A listing of some of these follows. This is part of what is, no doubt, an accumulating body of qualitative observations.

- *Information and Orientation:* It is quite rare for patient groups (5% or fewer) to not know the current U.S. president's name. It suggests the possibility of pronounced adaptational and cognitive deficits (J. Ryan & Paul, 1999).
- *Verbal Paired Associates:* Although the WMS-III norms do not take into account sex differences, females typically perform better than males on Verbal Paired Associates (i.e., $M = 10.58$ for females versus 8.46 for males on total recall scales scores; Basso, Harrington, Matson, & Lowery, 2000). This effect is moderately strong (approximately 3 subscale points) and should, therefore, be considered when making interpretations.

- *Logical Memory I and II:* Excessive embellishment of stories may be a maneuver to compensate for or cover up difficulty remembering accurate information. This may result in coherent elaboration or more illogical confabulations. A further behavioral observation is to note whether a client remembers primarily the global gist of the story as opposed to quite specific linear details. This may suggest either a global, holistic mode of processing as opposed to a more linear approach.
- *Letter-Number Sequencing:* Because letter-number sequencing is quite a demanding task, this subtest is the most likely to produce stimulus overload for the patient. The person might look frustrated or say that the task is too difficult ("You're expecting too much of me"). This suggests that he or she might experience similar frustration in everyday situations that similarly require complex reasoning (multitasking).

4. Evaluate Whether Change Has Occurred (relevant only for repeat testing)

Sometimes Wechsler memory scale scores are used to document deterioration or to monitor improvement. It is tempting to peruse pretest and posttest scores and quickly infer that some sort of actual change has occurred in the patient's level of functioning. For example, a client might have had a WMS General Memory Index score of 80 directly after a head injury and, three months later, achieved a score of 85. It might, therefore, be inferred that the patient's memory has improved. However, this does not take into consideration factors such as practice effects, regression to the mean, or the relative reliability of the measure. The improvement between the pretest of 80 and the posttest of 85 might simply be the result of the patient's practicing the tasks three months previously, or the difference might simply be measurement error (reflected in its test-retest reliability). To provide a more reliable measure of change, Iverson (1999) has calculated the following change indexes:

Auditory Immediate Index	11
Visual Immediate Index	10
Immediate Memory Index	10
Auditory Delayed Index	13
Visual Delayed Index	10
Auditory Recognition Delayed Index	15
General Memory Index	8
Working Memory Index	9

To be at least 80% certain that clinically meaningful change has occurred, a patient must have a difference of equal to (or greater than) the values indicated in the right column. For example, a patient should have gone from a General Memory Index score of 80 to at least 88 (an increase of 8 points) to be 80% certain that actual change had occurred. However, it should be stressed that these values were derived from patients with traumatic brain injury; therefore, these values may not necessarily transfer to other patient groups. The values have also been rounded to the nearest whole number. Finally, the previous statistical evaluation of change accounts for the unreliability of

the instrument, which does not necessarily prove that the personal or social significance of the change in scores has been demonstrated (see Beutler & Moleiro, 2001). Determining the personal and clinical meaning of changed scores requires clinicians to integrate information from a wider variety of sources.

RECOMMENDED READING

Franzen, M. D., & Iverson, G. L. (2000). The Wechsler Memory Scales. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration.* New York: Wiley.

Lichtenberger, E. O., Kaufman, A. S., & Lai, Z. C. (2002). *Essentials of WMS-III assessment.* New York: Wiley.

Chapter 7

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

The Minnesota Multiphasic Personality Inventory (MMPI)* is a standardized questionnaire that elicits a wide range of self-descriptions scored to give a quantitative measurement of an individual's level of emotional adjustment and attitude toward test taking. Since its original development by Hathaway and McKinley in 1940, the MMPI has become the most widely used clinical personality inventory, with more than 10,000 published research references (Boccaccini & Brodsky, 1999; Camara et al., 2000; C. Piotrowski, 1999; Watkins et al., 1995). Thus, in addition to its clinical usefulness, the MMPI has stimulated a vast amount of literature.

The 1943 MMPI test format consisted of 504 affirmative statements that could be answered "True" or "False." The number of items was later increased to 566 through the inclusion of repeat items and Scales 5 (Masculinity-Femininity) and 0 (Social Introversion). The 1989 restandardization retained the same basic format but altered, deleted, and/or added a number of items, which resulted in a total of 567. The different categories of responses can be either hand or computer scored and summarized on a profile sheet. An individual's score as represented on the profile form can then be compared with the scores derived from different normative samples.

The original MMPI had 13 standard scales, of which 3 related to validity and 10 related to clinical or personality indices. The more recent MMPI-2 and MMPI-A have maintained the original 10 clinical/personality scales as well as the original 3 validity scales, but the total number of validity scales has been increased (see Table 7.1). The clinical and personality scales are known both by their scale numbers and by scale abbreviations. Additional options are available to refine the meaning of the clinical scales as well as provide additional information. These include scales based on item content (content scales), subscales for the clinical and personality scales based on clusters of content-related items (Harris-Lingoes subscales), assessment of items and item clusters that relate to relevant dimensions (critical items) and empirically derived new scales (supplementary scales). New scales are still being researched and reported in the literature. The result of these developments is an extremely diverse and potentially useful test that can be interpreted, refined, and expanded from a variety of different perspectives.

*MMPI-2™ (Minnesota Multiphasic Personality Inventory-2)™ Test Booklet. Copyright © 1942, 1943 (renewed 1970), 1989 by the Regents of the University of Minnesota. All rights reserved. Used by permission of the University of Minnesota Press. "MMPI-2" and "Minnesota Multiphasic Personality-2" are trademarks owned by the Regents of the University of Minnesota.

Table 7.1 Validity, Basic (Clinical), and Content Minnesota Multiphasic Personality Inventory Scales

Name	Abbreviation	Scale No.	No. of Items
Validity scales			
Cannot say	?		
Variable response inconsistency	VRIN		98
True response inconsistency	TRIN		40
Lie	*L*		15
Infrequency	*F*		60
Correction	*K*		30
F back	*F(b)*		40
Basic (clinical) scales			
Hypochondriasis	*Hs*	1	32
Depression	*D*	2	57
Hysteria	*Hy*	3	60
Psychopathic deviant	*Pd*	4	50
Masculinity-femininity	*Mf*	5	56
Paranoia	*Pa*	6	40
Psychasthenia	*Pt*	7	48
Schizophrenia	*Sc*	8	78
Hypomania	*Ma*	9	46
Social introversion	*Si*	0	69
Content scales			
Anxiety	ANX		23
Fears	FRS		23
Obsessiveness	OBS		16
Depression	DPS		33
Health concerns	HEA		36
Bizarre mentation	BIZ		23
Anger	ANG		16
Cynicism	CYN		23
Antisocial practices	ASP		22
Type A	TPA		19
Low self-esteem	LSE		24
Social discomfort	SOD		24
Family problems	FAM		25
Work interference	WRK		33
Negative treatment indicators	TRT		26

The contents for the majority of MMPI questions are relatively obvious and deal largely with psychiatric, psychological, neurological, or physical symptoms. However, some of the questions are psychologically obscure because the underlying psychological process they are assessing is not intuitively obvious. For example, item 68, "I sometimes tease animals" is empirically answered "False" more frequently by depressed subjects than normals. Thus, it was included under Scale 2 (Depression) even though it does not, on the surface, appear to directly assess an individual's degree of depression.

For the most part, however, the statements are more direct and self-evident, such as item 56, "I wish I could be as happy as others seem to be" (True) or 146, "I cry easily" (True), both of which also reflect an examinee's level of depression. The overall item content is extremely varied and relates to areas such as general health, occupational interests, preoccupations, morale, phobias, and educational problems.

After a test profile has been developed, the scores are frequently arranged or coded in a way that summarizes and highlights significant peaks and valleys. However, to accurately interpret the test, both the overall configuration of the different scales and the relevant demographic characteristics of the client must be taken into consideration. In many instances, the same scaled score on one test profile can mean something quite different on another person's profile when the elevations or lowerings of other scales are also considered. For example, an elevated Scale 3 (Hysteria) may indicate an individual who denies conflict, demands support from others, expresses optimism, and is somewhat interpersonally naive. However, if this elevation is also accompanied by a high 4 (Psychopathic Deviate), there is likely to be a strong undercurrent of repressed anger. This anger is usually expressed indirectly, and any negative effects on others are likely to be strongly denied. Thus, it is important for the clinician to avoid the use of purely quantitative or mechanical formulas for interpreting the profile and instead examine the scores in the overall context of the other scale elevations and lowerings. Not only should a particular scale be examined in the context of the overall test configuration, but also additional sources such as demographic characteristics (age, education, socioeconomic status, ethnicity), behavioral observations, other psychometric devices, and relevant history can often increase the accuracy, richness, and sensitivity of personality descriptions.

A further important, general interpretive consideration is that the scales represent measures of personality traits rather than simply diagnostic categories. Although the scales were originally designed to differentiate normal from abnormal behavior, it is generally regarded as far more useful to consider that the scales indicate clusters of personality variables. For example, Scale 2 (Depression) may suggest characteristics such as mental apathy, self-deprecation, and a tendency to worry over even relatively small matters. This approach characterizes the extensive research performed on the meanings of the two highest scales (two-point code types), which are summarized later in this chapter. Rather than merely labeling a person, this descriptive approach creates a richer, more in-depth, and wider assessment of the individual who is being tested.

HISTORY AND DEVELOPMENT

The original development of the MMPI was begun in 1939 at the University of Minnesota by Starke R. Hathaway and J. Charnley McKinley. They wanted an instrument that could serve as an aid in assessing adult patients during routine psychiatric case workups and that could accurately determine the severity of their disturbances. Furthermore, Hathaway and McKinley were interested in developing an objective estimate of the change produced by psychotherapy or other variables in a patient's life.

The most important approach taken during construction of the MMPI was empirical criterion keying. This refers to the development, selection, and scoring of items within the scales based on some external criterion of reference. Thus, if a clinical population

was given a series of questions to answer, the individuals developing the test would select questions for inclusion or exclusion based on whether this clinical population answered differently from a comparison group. Even though a theoretical approach might be used initially to develop test questions, the final inclusion of questions would not be based on this theoretical criterion. Instead, test questions would be selected based on whether they were answered in a direction different from a contrasted group. For example, a test constructor may believe that an item such as "Sometimes I find it almost impossible to get up in the morning" is a theoretically sound statement to use in assessing depression. However, if a sample population of depressed patients did not respond to that question differently from a normative group, the item would not be included. Thus, if a person with hysterical traits answers "True" to the statement "I have stomach pains," whether he or she actually does have stomach pains is less important—from an empirical point of view—than the fact that the individual *says* he or she does. In other words, the final criterion for inclusion of items in an inventory is based on whether these items are responded to in a significantly different manner by a specified population sample.

Using this method, Hathaway and McKinley began with an original item pool of more than 1,000 statements derived from a variety of different sources, including previously developed scales of personal and social attitudes, clinical reports, case histories, psychiatric interviewing manuals, and personal clinical experience. Of the original 1,000 statements, many were eliminated or modified. The result was 504 statements that were considered to be clear, readable, not duplicated, and balanced between positive and negative wording. The statements themselves were extremely varied and were purposely designed to tap as wide a number of areas in an individual's life as possible. The next step was to select different groups of normal and psychiatric patients to whom the 504 questions could be administered. The normals were primarily friends and relatives of patients at the University of Minnesota hospitals who were willing to complete the inventory. They consisted of 226 males and 315 females, who were screened with several background questions about age, education, marital status, occupation, residence, and current medical status. Individuals who were under the care of a physician at the time of the screening were excluded from the study. This group was further augmented by the inclusion of other normal subjects, such as recent high school graduates, Work Progress Administration workers, and medical patients at the University of Minnesota hospitals. This composite sample of 724 individuals was closely representative in terms of age, sex, and marital status of a typical group of individuals from the Minnesota population, as reflected in the 1930 census. The clinical group comprised patients who represented the major psychiatric categories being treated at the University of Minnesota hospitals. These patients were divided into clear subgroups of approximately 50 in each category of diagnosis. If a patient's diagnosis was at all in question, or if a person had a multiple diagnosis, he or she was excluded from the study. The resulting subgroups were hypochondriasis, depression, hysteria, psychopathic deviate, paranoia, psychasthenia, schizophrenia, and hypomania.

After the normals and psychiatric patients had been administered the 504-item scale, Hathaway and McKinley could then compare their responses. Each item that correctly differentiated between these two groups was included in the resulting clinical scale. For example, item 40, "Much of the time my head seems to hurt all over," was

answered "True" by 12% of the sample of hypochondriacs and by only 4% of the normals. It was thus included in the clinical scale for hypochondriasis. The comparisons, then, were between each clinical group and the group of normals rather than among the different clinical groups themselves. This was the selection procedure used to develop tentative clinical scales.

Still another step was included in the scale constructions. The fact that an item was endorsed differently by the group of 724 Minnesota normals than by the patients from various clinical populations did not necessarily indicate that it could be used successfully for clinical screening purposes. Thus, an attempt was made to cross-validate the scales by selecting a new group of normals and comparing their responses with a different group of clinical patients. The items that still provided significant differences between these groups were selected for the final version of the scales. It was reasoned, then, that these items and the scales comprising these items would be valid for differential diagnosis in actual clinical settings.

Whereas this procedure describes how the original clinical scales were developed, two additional scales that used slightly different approaches were also included. Scale 5 (Masculinity-Femininity) was originally intended to differentiate male homosexuals from males with a more exclusively heterosexual orientation. However, few items were found that could effectively perform this function. The scale was then expanded to distinguish items that were characteristically endorsed in a certain direction by the majority of males from those that were characteristically endorsed in a certain direction by females. This was accomplished in part by the inclusion of items from the Terman and Miles I Scale (1936). The second additional scale, Social Introversion (*Si*), was developed by Drake in 1946. It was initially developed by using empirical criterion keying in an attempt to differentiate female college students who participated extensively in social and extracurricular activities from those who rarely participated. It was later generalized to reflect the relative degree of introversion for both males and females.

It soon became apparent to the test constructors that persons could alter the impression they made on the test because of various test-taking attitudes. Hathaway and McKinley thus began to develop several scales that could detect the types and magnitude of the different test-taking attitudes most likely to invalidate the other clinical scales. Four scales were developed: the Cannot say (?), the Lie (*L*), the Infrequency (*F*), and the Correction (*K*). The Cannot say scale (?) is simply the total number of unanswered questions. If a high number of these are present, it would obviously serve to reduce the validity of the overall profile. High scores on the Lie scale indicate a naive and unsophisticated effort on the part of the examinee to create an overly favorable impression. The items selected for this scale were those that indicated a reluctance to admit to even minor personal shortcomings. The *F* scale is composed of those items endorsed by fewer than 10% of normals. A high number of scorable items on the *F* scale, then, reflects that the examinee is endorsing a high number of unusually deviant responses.

K, which reflects an examinee's degree of psychological defensiveness, is perhaps the most sophisticated of the validity scales. The items for this scale were selected by comparing the responses of known psychiatric patients who still produced normal MMPIs (clinically defensive) with "true" normals who also produced normal MMPIs. Those items that differentiated between these two groups were used for the *K* scale. Somewhat later, the relative number of items endorsed on the *K* scale was used as a "correction"

factor. The reasoning behind this was that, if some of the scales were lowered because of a defensive test-taking attitude, a measure of the degree of defensiveness could be added into the scale to compensate for this. The result would theoretically be a more accurate appraisal of the person's clinical behavior. The scales that are not given a *K* correction are those whose raw scores still produced an accurate description of the person's actual behavior. However, there have been some questions regarding the effectiveness of the *K* correction in some settings. As a result, clinicians have the choice of whether they wish to use MMPI-2 profile sheets with or without the *K* correction, and the MMPI-A has omitted the use of the *K* correction altogether.

Since the publication of the original MMPI, special scales and numerous adjunctive approaches to interpretation have been developed. A primary strategy has been content interpretation. The most frequently used are the Harris and Lingoes subscales, Wiggins Content Scales, and several different listings of critical items, which can potentially provide important qualitative information regarding an examinee. In addition, many supplementary scales have been developed, such as the Anxiety Scale, the MacAndrew Scale to assess the potential for substance abuse, and the Ego Strength Scale to estimate the extent to which a person will benefit from insight-oriented therapy. Each of these approaches can be used as an adjunct in interpreting the traditional clinical scales and/or experimental scales for assessing or researching specific populations (see Butcher, Graham, Williams, & Ben-Porath, 1990; J. Graham, 2000; C. Williams, Butcher, Ben-Porath, & Graham, 1992).

In addition to innovations in scales and interpretations, the MMPI has been used in a wide number of settings for extremely diverse areas. Most studies have focused on the identification of medical and psychiatric disorders as well as on uses in forensic contexts (Deardorff, 2000; Greene, 2000; K. Pope, Butcher, & Seelen, 2000), and on expanding or further understanding the psychometric properties of the MMPI. Other frequent topics include alcoholism, aging, locus of control, computer-based interpretation, chronic pain, and the assessment of different occupational groups. The MMPI has been translated into a number of different languages and has been used in a wide range of different cross-cultural contexts (see Butcher, 1996; Cheung & Ho, 1997; Greene, 1991; G. Hall, Bansal, & Lopez, 1999; Handel & Ben-Porath, 2000).

Criticisms of the original MMPI have primarily centered on its growing obsolescence, difficulties with the original scale construction, inadequacy of its standardization sample, and difficulties with many of the items (Butcher & Pope, 1989; Helmes & Reddon, 1993). Problems with the items included sexist wording, possible racial bias, archaic phrases, and objectionable content. In addition, the original norms had poor representation of minorities and were inappropriate in making comparisons with current test takers. Further problems have related to inconsistent meanings associated with *T*-score transformations.

These criticisms led to an extensive restandardization of the MMPI, which began in 1982. Despite the need to make major changes, the restandardization committee wanted to keep the basic format and intent of the MMPI as intact as possible so that the extensive research base collected over the past 50 years would still be applicable to the restandardized version. As a result, the following six goals were established (Butcher & Pope, 1989):

1. The deletion of obsolete or objectionable items.
2. Continuation of the original validity and clinical scales.
3. The development of a wide, representative normative sample.
4. Norms that would most accurately reflect clinical problems and would result in a uniform percentile classification.
5. The collection of new clinical data that could be used in evaluating the items and scales.
6. The development of new scales.

The restandardization used a special research form consisting of the original 550 items (of which 82 were modified) and additional 154 provisional items used for the development of new scales. Even though 82 of the original items were reworded, their psychometric properties apparently were not altered (Ben-Porath & Butcher, 1989). The resulting 704-item form (Form AX) was administered to 1,138 males and 1,462 females from seven different states, several military bases, and a Native American reservation. The subjects were between the ages of 18 and 90 and were contacted by requests through direct mail, advertisements in the media, and special appeals. The resulting restandardization sample was highly similar to the 1980 U.S. census in almost all areas with the exception that they were somewhat better educated than the overall population.

The MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) differs from the older test in a number of ways. The *T* scores that subjects obtain are generally not as deviant as those from the earlier version. In addition, the *T* scores were designed to produce the same range and distribution throughout the traditional clinical scales (except for Scales 5 and 0). The practical result is that *T* scores of 65 or greater are considered to be in the clinical range (versus a cutoff score of 70 for the MMPI). Also, the percentile distributions are uniform throughout the different clinical scales (whereas they were unequal for the MMPI). The test booklet itself contains 567 items, but the order has been changed so that the traditional scales (3 validity and 10 clinical) can be derived from the first 370 items. The remaining 197 items (371 to 567) provide different supplementary, content, and research measures. A number of new scales were included along with new, subtle, adjunctive measures of test validity, separate measures of masculinity and femininity, and 15 additional content scales measuring specific personality factors (Anxiety, Health concerns, Cynicism, etc.). An extensive research base has accumulated related to areas such as the validity of MMPI/MMPI-2 code types, use with special populations, the ability to distinguish over- or underreporting of symptoms, and comparability between the original MMPI, MMPI-2, and MMPI-A.

Early on, it was noticed that the original MMPI produced different scale elevations for adolescents than for adults. This resulted in the development of different sets of recommended norms for use with adolescent populations (Archer, 1987; Colligan & Offord, 1989; Klinefelter, Pancoast, Archer, & Pruitt, 1990; Marks, Seeman, & Haller, 1974). However, many practitioners and researchers felt that, even with the use of adolescent norms, there were still considerable difficulties. Specifically, it was too long, the reading level was too high, there was a need for contemporary norms, more of the

content needed to assess problems specifically related to adolescents, and some of the language was outmoded and/or inappropriate (Archer, Maruish, Imhof, & Piotrowski, 1991). In response to these issues, the restandardization committee for the MMPI-2 decided in 1989 to develop the MMPI-Adolescent (MMPI-A), which was first made available in 1992 (Butcher et al., 1992). It was normed against a generally representative group of 805 males and 815 females between the ages of 14 and 18. The main discrepancy between the normative group and comparison with U.S. census data was that the parents of the normative group were better educated. Despite the similarity with the MMPI and MMPI-2, there are several important differences. Fifty-eight items were deleted from the original standard scales, some of the wording of items was changed, and new items relevant to adolescent concerns were included. The result has been the inclusion of four new validity scales (VRIN, TRIN, *F*1, *F*2) in addition to the earlier validity scales (*L, F, K*). There are also six supplementary scales (e.g., Immaturity Scale, Anxiety, Repression) and additional newly developed content scales (e.g., A-dep/Adolescent Depression). To counter claims that the MMPI is too long, especially for adolescents, the new MMPI-A contains 478 items, thereby shortening the administration time. This can be shortened even further by administering only the first 350 items, still sufficient to obtain the validity and standard clinical scales. Thus, the MMPI-A is both strongly related to the MMPI and MMPI-2 (and their respective databases) but also has a number of important distinctive features of its own.

RELIABILITY AND VALIDITY

Reliability studies on the original MMPI indicate that it had moderate levels of temporal stability and internal consistency. For example, Hunsley, Hanson, and Parker (1988) completed a meta-analysis of studies performed on the MMPI between 1970 and 1981, and concluded, "all MMPI scales are quite reliable, with values that range from a low of .71 (Scale *Ma*) to a high of .84 (Scale *Pt*)" (p. 45). Their analysis was derived from studies that included a wide range of populations, intervals that ranged from one day to two years, and a combined sample size exceeding 5,000. In contrast to Hunsley et al., some authors have reported that the fluctuations in some of the scales are sufficiently wide to question their reliabilities (Hathaway & Monachesi, 1963; Mauger, 1972). Proponents of the MMPI counter that some fluctuation in test scores are to be expected. This is especially true for psychiatric populations because the effects of treatment or stabilization in a temporary crisis are likely to be reflected in a patient's test performance (J. Graham, Smith, & Schwartz, 1986). Bergin (1971) has demonstrated that Scale 2 (Depression) is particularly likely to be lowered after successful treatment. Similarly, Scale 7 (Psychasthenia) would be likely to alter according to a person's external situation. Thus, test-retest reliability may actually be an inappropriate method of evaluating these scales for certain populations. This defense of the test's reliability is somewhat undermined by the observation that test-retest reliability is actually slightly more stable for psychiatric populations than for normals. Whereas the median range for psychiatric patients is about .80, median reliabilities for normals are about .70. Split-half reliabilities are likewise moderate, having an extremely wide range from .05 to .96, with median correlations in the .70s (Hunsley et al., 1988).

Reliability reported in the MMPI-2 manual indicates moderate test-retest reliabilities. However, test-retest reliabilities were calculated for a narrow population over short-term retesting intervals. Reliabilities for normal males over an average interval of 8.58 days (*Mdn* = 57 days) ranged from a low of .67 for Scale 6 to a high of .92 for Scale 0 (Butcher et al., 1989). A parallel sample of females over the same retesting interval produced similar reliabilities ranging from .58 (Scale 6) to .91 (Scale 0). Standard error of measurements for the different scales ranged from 2 to 3 raw score points (Butcher et al., 1989; Munley, 1991). Future studies will no doubt provide a further evaluation of the MMPI-2's reliability over longer intervals and for various population groups.

One difficulty with the MMPI/MMPI-2 lies in the construction of the scales themselves. The intercorrelations between many of the scales are quite high, which results primarily from the extensive degree of item overlap. Sometimes, the same item will be simultaneously used for the scoring of several different scales, and most of the scales have a relatively high proportion of items common to other scales. For example, Scales 7 (Psychasthenia) and 8 (Schizophrenia) have high overlap, which is reflected in correlations ranging from .64 to .87, depending on the population sampled (Butcher et al., 1989; Dahlstrom & Welsh, 1960). Scale 8, which has the highest number of items (78), has only 16 items that are unique to it (Dahlstrom, Welsh, & Dahlstrom, 1972). Similarly, Scale *F* (Infrequency) is highly correlated with Scales 7 (*Pt*), 8 (*Sc*), and the Bizarre Mentation content scale. The practical implication is that interpreters need to be quite cautious about inferring a "fake bad" profile if profile *F* is elevated along with 7 (*Pt*), 8 (*Sc*), and Bizarre Mentation. Several factor analytic studies have been conducted that were motivated in part by a need to further understand the high intercorrelations among scales. These studies have not found any consistent numbers and types of factors. The numbers of factors range between 2 (Dahlstrom et al., 1972; Dahlstrom, Welsh, & Dahlstrom, 1975; D. Jackson, Fraboni, & Helms, 1997) and 9 (Archer & Krishnamurthy, 1997a; Costa, Zonderman, Williams, & McCrae, 1985) and even up to 21 (J. H. Johnson, Null, Butcher, & Johnson, 1984). This suggests that these factors are not highly differentiated.

The different scales correlate so highly, in part, because the original selection of the items for inclusion in each scale was based on a comparison of normals with different clinical groups. The items, then, were selected based on their differentiation of normals from various psychiatric populations, rather than on their differentiation of one psychiatric population from another. Although the psychiatric groups varied from the normals on several traits, this manner of scale construction did not develop accurate measurements of these different traits. Rather, the scales are filled with many heterogeneous items and measure multidimensional, often poorly defined attributes. This approach has also led to many items being shared with other scales. In contrast, an approach in which specific psychiatric groups had been compared with one another would have been more likely to have resulted in scales with less item overlap and with the ability to measure more unidimensional traits.

A partial defense of item overlap is that for complex, multidimensional variables such as pathological syndromes, important relationships would be expected with other similar constructs. If these other constructs were being measured on the same test, it would further be expected that there would be scale overlap on these theoretically and clinically

related syndromes (Dahlstrom et al., 1972). For example, depression is a common feature among several categories of psychopathology. Thus, it would be theoretically related to conditions such as hypochondriasis, schizophrenia, and anxiety. This in turn would result in expected intercorrelations between scales, and would produce scales that, while intercorrelated, would still have subtle and clinically different meanings (Broughton, 1984). Thus, the multidimensionality of the scales combined with their item overlap would be not so much a weaknesses of the MMPI/MMPI-2/MMPI-A, but would be expected, given the nature of the constructs. Accurate interpretation, however, would need to include an awareness of the subtle differences and similarities between scales.

An issue related to MMPI/MMPI-2/MMPI-A scale multidimensionality is that elevations can often occur for a variety of reasons. For example, an elevation on 4 (Psychopathic Deviance) might result from family discord, poor peer relations, alienation from self and society, and/or acting out associated with legal difficulties. This means that a person interpreting an elevated Scale 4 (Psychopathic Deviance) might potentially infer antisocial acting out when family discord is the major reason for the scale elevation. To enhance the likelihood of accurate interpretations, practitioners need to carefully evaluate the meanings of scale elevations. This might include looking at the content of selected items (critical items), scoring the Harris-Lingoes subscales, considering the meanings of content or supplementary scales, referring to published MMPI research, and integrating the results from the client's history and relevant behavioral observations. Differentiating which of these scale dimensions is most relevant can be quite challenging for the practitioner.

A further difficulty relating to scale construction is the imbalance in the number of true and false items. In the *L* scale, all the items are scorable if answered "False"; on the *K* scale, 29 of 30 items are scored if answered "False"; and Scales 7, 8, and 9 have a ratio of approximately 3 to 1 of true compared with false items. The danger of this imbalance is that persons having response styles of either acquiescing ("yea-saying") or disagreeing ("nay-saying") may answer according to their response style rather than to the content of the items. A more theoretically sound approach to item construction would have been to include an even balance between the number of true and false answers. Some authors (A. Edwards, 1957, 1964; D. Jackson et al., 1997) have even suggested that test results do not reflect psychological traits as much as generalized test-taking attitudes. Thus, a controversy has arisen over "content variance," in which an examinee is responding to the content of the items in a manner that will reflect psychological traits rather than "response style variance," in which responses reflect more the examinee's tendency to respond in a certain biased direction. In a review of the literature, Koss (1979) concluded that, although response sets can and do exist, the examinee's tendency to respond accurately to the item content is far stronger. The MMPI-2 restandardization committee has also developed the Variable Response Inconsistency (VRIN) and True Response Inconsistency (TRIN) scales to help detect invalid profiles caused by inconsistent or contradictory responding. These scales have been specifically designed to detect either response acquiescence or response nonacquiescence and thus should help counter the potential complications resulting from imbalanced keying.

The difficulties associated with reliability and scale construction have led to challenges to the MMPI's validity. Rodgers (1972) has even referred to the MMPI as a "psychometric nightmare." However, although the strict psychometric properties present

difficulties, this has been somewhat compensated by extensive validity studies. More specifically, the meanings of two- and three-point profile code types have been extensively researched, as have the contributions that the MMPI can make toward assessing and predicting specific problem areas. Dahlstrom et al. (1975), in Volume 2 of their revised MMPI handbook, cited 6,000 studies investigating profile patterns. This number is continually increasing (see, e.g., Butcher 2000; DuAlba & Scott, 1993; Gallucci, 1994; J. Graham, Ben-Porath, & McNulty, 1999; McNulty, Ben-Porath, & Graham, 1998), and past studies provide extensive evidence of the MMPI's construct validity. For example, elevations on Scales 4 (*Pd*) and 9 (*Ma*) have been associated with measures of impulsivity, aggression, substance abuse, and sensation seeking among adolescent inpatients (Gallucci, 1994). In addition, the degree to which individuals improve from psychotherapy was predicted based on elevations on the content scales of Anxiety (ANX) and Depression (DEP; Chisholm, Crowther, & Ben-Porath, 1997). Finally, high scores on Scale 0 (*Si*) have been associated with persons who have low self-esteem, social anxiety, and low sociability (Sieber & Meyers, 1992). Individual clinicians can consult research on code types to obtain specific personality descriptions and learn of potential problems to which a client may be susceptible. The extensiveness and strength of these validity studies have usually been regarded as major assets of the MMPI and are important reasons for its continued popularity.

In addition to studying the correlates of code type, another approach to establishing validity is to assess the accuracy of inferences based on the MMPI. Early studies by Kostlan (1954) and Little and Shneidman (1959) indicated that the MMPI is relatively more accurate than other standard assessment instruments, especially when the MMPI was combined with social case history data. This incremental validity of the MMPI has been supported in later reviews by Garb (1998b) and J. Graham and Lilly (1984). For example, the accuracy of neurologists' diagnoses was found to increase when they added an MMPI to their patient data (S. Schwartz & Wiedel, 1981). Garb (1998b) concluded that the MMPI was more accurate than social history alone, was superior to projectives, and that the highest incremental validity was obtained when the MMPI was combined with social history. In addition, incremental validity of the new MMPI-2 content scales has been found in that they both expanded on and increased the validity of the standard clinical scales (Barthlow, Graham, Ben-Porath, & McNulty, 1999; Ben-Porath, McCully, & Almagor, 1993).

ASSETS AND LIMITATIONS

The previous discussion on reliability and validity highlights several issues associated with the MMPI. These include moderate levels of reliability, extensive length, and problems related to the construction of the scales, such as item overlap, high intercorrelations among scales, and multidimensional poorly defined variables. Some of the criticisms of the original MMPI relating to obsolete norms, offensive items, and poorly worded items have been largely corrected with the publication of the MMPI-2 and MMPI-A. The MMPI also has a number of strengths, along with other weaknesses.

One caution stemming from the construction of the original MMPI is that it generally does not provide much information related to normal populations. The items were

selected on the basis of their ability to differentiate a bimodal population of normals from psychiatric patients. Thus, extreme scores can be interpreted with a high degree of confidence, but moderate elevations must be interpreted with appropriate caution. An elevation in the range of one standard deviation above the mean is more likely to represent an insignificant fluctuation of a normal population than would be the case if a normally distributed group had been used for the scale construction. This is in contrast to a test such as the California Personality Inventory (CPI), which used a more evenly distributed sample (as opposed to a bimodal one) and, as a result, can make meaningful interpretations based on moderate elevations. The MMPI-2 partially addresses this difficulty as it has used broad contemporary norms for its comparisons, combined with uniform *T* scores (Tellegen & Ben-Porath, 1992). However, evaluation of normals can be complicated by the observation that normal persons sometimes achieve high scores. Despite these difficulties, the use and understanding of nonclinical populations have been increasing (J. Graham & McCord, 1985; Keiller & Graham, 1993). In particular, uses have included screening personnel for sensitive jobs such as air traffic controllers, police officers, and nuclear plant operators.

Although there have been a number of notable improvements with the MMPI-2, issues have been raised regarding comparability between the two versions. In defense of their comparability are the many similarities in format, scale descriptions, and items. In particular, Ben-Porath and Butcher (1989) found that the effects of rewriting 82 of the MMPI items for inclusion in the MMPI-2 were minimal. The rewritten items had no effect on any of the validity, clinical, or special scales when comparisons were made between administrations of the original and restandardized versions using college students. This provided some support for Butcher and Pope's (1989) contention that the MMPI-2 validity and clinical scales measure "exactly what they have always measured" (p. 11). Further studies have generally found that there are few differences based on individual scale comparisons (Ben-Porath & Butcher, 1989; Chojnacki & Walsh, 1992; Harrell, Honaker, & Parnell, 1992; L. Ward, 1991). Similarly, number of elevated scales between the two forms does not seem to be significantly different, and there has been 75% agreement regarding whether a subject's profile was considered to be within normal limits (Ben-Porath & Butcher, 1989).

Despite these similarities, the use of the restandardization norms and the use of uniform *T* scores have created differences in two-point codes among different population samples, including differences among 31% of the code types derived from general psychiatric patients (Butcher et al., 1989), 22% of peace officers (Hargrave, Hiatt, Ogard, & Karr, 1994), 39% to 42% of psychiatric inpatients (D. Edwards, Morrison, & Weissman, 1993; H. Weissman, 1992), and a full 50% of both university students (H. Weissman, 1992) and forensic populations (Humphrey & Dahlstrom, 1995). The greatest level of disagreements was for poorly defined code types (mild to moderate elevations combined with more than two "competing" scales). In contrast, well-defined code types (highly elevated and without "competing" third or fourth most elevated scales) had considerably higher concordance (Tellegen & Ben-Porath, 1993). This suggests that special care should be taken regarding poorly defined code types, and, if more than two scales are elevated, the meanings of the relatively high scales not included in the code should be given particular interpretive attention.

These discrepancies in code types seem to question the exact transferability of past code type research on the MMPI onto the more recent MMPI-2 (and MMPI-A). However, the most important question is the extent to which the MMPI-2 accurately describes an individual's relevant behaviors. The research that has been done on the MMPI-2 does support the conclusion that scores on the MMPI-2 predict the same sorts of behaviors that were found with the earlier MMPI. (Archer, Griffin, & Aiduk, 1995; J. Graham et al., 1999; Timbrook & Graham, 1994). As research continues to explore the MMPI-2 (and MMPI-A) validity, it will progressively mean that interpretations based on these newer versions can rely on their own research base rather than having to depend on the earlier work done with the MMPI.

As highlighted in the previous section, a traditional asset of the MMPI/MMPI-2/MMPI-A has been extensive and ongoing code type studies. However, difficulties with these studies have been recently noted. First, some studies have tried to be extremely inclusive in deciding which codes to evaluate. In contrast, others have been quite restrictive (i.e., including only clearly defined code types). Inclusion/exclusion among the different studies has ranged from 24 to 99% (McGrath & Ingersoll, 1999a). The practical implication for clinicians is considering the degree to which their code type classifications parallel those of research. If specific clinicians are highly inclusive about what they consider to be interpretable code types, they may place unwarranted faith in their interpretations if the body of research they are drawing from has used quite restrictive criteria (i.e., J. Graham et al., 1999 used only well-defined code types). A further concern is that the mean effect size across studies was quite variable, with a high of .74 and low of .02 (McGrath & Ingersoll, 1999b; G. Meyer & Archer, 2001). In addition, effect sizes were found to vary among different scales and code types. Therefore, practitioners may not only be placing unwarranted faith in some of their interpretations, but also the validity of the interpretations they do make are likely to vary according to which scale/code type they are interpreting.

In all versions of the MMPI, the scale labels can be misleading because they use traditional diagnostic categories. A person might read a scale such as Schizophrenia and infer that a person with a peak on that scale, therefore, fits the diagnosis of schizophrenia. Although it was originally hoped that the MMPI could be used to make differential psychiatric diagnoses, it was soon found that it could not adequately perform this function. Thus, even though schizophrenics score high on Scale 8, so do other psychotic and nonpsychotic groups. Also, moderate elevations can occur for some normal persons. With the publication of the third and fourth editions of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*; American Psychiatric Association, 1980, 1987, 1994), the traditional labels on which the scale names were based have become progressively more outdated. This causes further confusion related to diagnosis because the scales reflect older categories. For example, Scales 1, 2, and 3 are called the *neurotic triad,* and Scale 7 is labeled *Psychasthenia;* yet clinicians are often faced with the need to translate these outdated designations into *DSM-IV* (1994) terminology. This difficulty has been somewhat alleviated through research focusing on the frequencies of *DSM-III* (1980) and *DSM-III-R* (1987) classifications, which are related to different code types (Morey, Blashfield, Webb, & Jewell, 1988; Vincent et al., 1983). *DSM-III/DSM-IV* translations

have been further aided through the use of different content and supplementary scales that allow for broader descriptions of symptom patterns (Barthlow et al., 1999; Butcher et al., 1990; Levitt & Gotts, 1995; C. Williams et al., 1992).

To compensate for the difficulties related to scale labels, clinicians should become aware of the current meanings of the scales based on research rather than the meanings implied by the often misleading scale titles. This approach can be aided in part by using scale numbers rather than titles. For example, Scale 8 suggests attributes such as apathy, feelings of alienation, philosophical interests, poor family relations, and unusual thought processes rather than schizophrenia. It is the clinician's responsibility to determine which of these attributes are most characteristic of the person being evaluated. Clinicians should also be aware of the relationships among scales as represented by the extensive research performed on two- and three-point code types. Usually, the patterns or profiles of the scales are far more useful and valid than merely considering individual scale elevations. The extensiveness of research in this area represents what is probably the strongest asset of the MMPI. This volume of work has prevented the MMPI from becoming obsolete and has been instrumental in transforming it from a test of psychiatric classification into a far more wide-band personality inventory.

A further significant asset is the MMPI's immense popularity and familiarity within the field. Extensive research has been performed in a variety of areas, and new developments have included abbreviated forms, new scales, the use of critical items, an adolescent version, and computerized interpretation systems. The MMPI has been translated into more than 50 languages and is available in numerous countries. Normative and validity studies have been conducted on several different cultural groups (see Butcher, 1996; Handel & Ben-Porath, 2000), which makes possible the comparison of data collected from varying cultures. In contexts where no norms have been developed, at least the test format lends itself to the development of more appropriate norms that can then be used in these contexts.

A complicating aspect of the MMPI is that interpretations often need to take into account many demographic variables (Schinka, LaLone, & Greene, 1998). It has been demonstrated that age, sex, race, place of residence, intelligence, education, and socioeconomic status are all related to the MMPI scales. Often the same relative elevation of profiles can have quite different meanings when corrections are made for demographic variables. Some of the more important and well researched of these are discussed.

Age

Typically, elevations occur on Scales 1 and 3 for older normal populations (Leon, Gillum, Gillum, & Gouze, 1979). On the other hand, MMPI Scales *F,* 4, 6, 8, and 9 are commonly elevated for adolescent populations (Marks et al., 1974). These patterns have been accounted for in the MMPI-A by using separate norms. As the sampled population becomes older, the deviations of the latter group of scales tend to decrease. Further, Scale 9 is more commonly elevated in younger persons but decreases with age until it becomes the most frequent low point in older populations (Gynther & Shimkuras, 1966). As a rule, the left side of the profile (Scales 1, 2, and 3) increases with age, which parallels the trend in older persons toward greater concern with health (Scales 1 and 3) and depression (Scale 2). Conversely, the right side of the profile decreases with age, which

parallels a decrease in energy level (Scale 9), increased introversion (Scale 0 as well as 2), and decreased assertiveness (Scale 4). In specific cases, however, there may also be a complex interaction with gender, health, socioeconomic status, and ethnicity. In addition to considering scale elevations related to aging, it may be helpful to evaluate individual item content. Swensen, Pearson, and Osborne (1973) provide a list of 30 items that are likely to be affected by aging, such as MMPI item 9, "I am about as able to work as I ever was" (False) and MMPI item 261, "If I were an artist I would like to draw flowers" (True). An analysis of these items indicates that older persons generally express a decrease in hostility (MMPI items 39, 80, 109, 282, and 438), have more stereotypically "feminine" interests (MMPI items 132 and 261), and are more dutiful, placid, and cautious (Gynther, 1979).

As noted earlier, a significant feature of adolescent populations is a general elevation on many of the MMPI scales. This has led to considerable controversy over whether adolescents have more actual pathology (based on external behavioral correlates) or whether they merely have higher scores without correspondingly higher pathology (Archer, 1984, 1987, 1992a; Janus, Tolbert, Calestro, & Toepfer, 1996).The controversy has encouraged efforts to more clearly understand behavioral correlates of adolescent profiles (Archer & Jacobson, 1993; Basham, 1992; Janus et al., 1996; Spirito, Faust, Myers, & Bechtel, 1988). Most authors have encouraged the use of specific adolescent norms, such as those developed by Marks et al. (1974). A problem with the Marks et al. norms is that they may have been adequate for past assessment of adolescents (before and shortly after 1974), but current adolescent populations may require the use of more recently developed norms (Archer, Pancoast, & Klinefelter, 1989; Janus et al., 1996; Pancoast & Archer, 1988; C. Williams & Butcher, 1989a, 1989b), particularly the norms developed for the MMPI-A (Archer, 1992a; Butcher et al., 1992). A specific problem is that the older Marks et al. norms produce a high percentage of false negatives for contemporary older adolescents populations, and descriptors based on adult norms may actually be more accurate than descriptors based on the Marks et al. (1974) norms. Furthermore, Scale 5 (Masculinity-Femininity) does not seem to have external correlates for adolescents, and many of the more frequent adolescent code types have received only limited recent support (Williams & Butcher, 1989a, 1989b). Current researchers make the somewhat awkward recommendation that, when using the MMPI-2, both adolescent and adult descriptors should be used for developing interpretations of adolescent profiles (Archer, 1984, 1987). It was these and related issues that led to the development of the MMPI-A. The general consensus seems to be that using the MMPI-A (and norms based on them) results in behavioral descriptions that are at least as accurate as descriptors based on the MMPI/MMPI-2 (Archer, 1992a, 1992b; Butcher et al., 1992; Janus et al., 1996; Weed, Butcher, & Williams, 1994).

Ethnicity

The MMPI/MMPI-2 has been extensively studied to determine how appropriate it is to use with culturally divergent groups. This research has centered around both ethnically different (minority) groups within the United States as well as its use in different countries. There are a wide variety of possible reasons why persons from different

cultural groups might score in a certain direction. Although scores may be due to the accurate measurement of different personality traits, they may also be the result of cultural tendencies to acquiesce by giving socially desirable responses, differing beliefs about modesty, role conflicts, or varying interpretations of the meaning of items. Profiles may also reflect the results of racial discrimination in that scales associated with anger, impulsiveness, and frustration may be elevated.

MMPI/MMPI-2 research on ethnic groups within the United States has centered on differences between African versus European Americans. Research on African American versus European Americans' MMPI performance has frequently indicated that African Americans are more likely to score higher on Scales *F,* 8, and 9 (Green & Kelley, 1988; Gynther & Green, 1980; Smith & Graham, 1981). This has resulted in considerable controversy over whether these differences indicate higher levels of actual pathology or merely reflect differences in perceptions and values without implying greater maladjustment. If the differences did not reflect greater actual pathology, then specialized subgroup norms would be required to correct for this source of error. However, reviews of over 30 years of research have concluded that, although African versus European American differences could be found for some populations, there was no consistent pattern to these differences across all populations (Greene, 1987, 1991; G. Hall, Bansal, & Lopez, 1999). What seemed of greater significance was the role of moderator variables, such as education, income, age, and type of pathology. When African American and European American psychiatric patients were compared according to level of education and type of pathology, their MMPI/MMPI-2 performances were the same (McNulty, Graham, Ben-Porath, & Stein, 1997; Timbrook & Graham, 1994). The issue of actual behavioral correlates of African American MMPI performance has generally not found differences between African American and European Americans. For example, ratings by clinicians (McNulty, Graham, Ben Porath, & Stein, 1997) and partners (Timbrook & Graham, 1994) were equally as accurate for both groups. In addition, the main behavioral features of 68/86 code types between African American and European Americans were the same (Clark & Miller, 1971). Furthermore, predictions based on African American and European American juvenile delinquents' MMPI scores were equally accurate for African Americans and European Americans (Green & Kelley, 1988; Timbrook & Graham, 1994). A final crucial finding has been that, even when mean differences have been found, they have been less than 5 *T*-score points difference (G. Hall et al., 1999; Stukenberg, Brady, & Klinetob, 2000). The magnitude of this difference is not clinically meaningful. Based on the preceding findings it would be premature to develop and use separate norms for African Americans. However, it would still be important for clinicians to continually be aware of any possible culturally relevant factors (i.e., effects of discrimination) which may cause unique elevations in an individual African American's profile.

Similar to African American versus European American comparisons, no consistent patterns have been found across different populations for Native American, Hispanics, and Asian Americans. For example, normal Native Americans scored higher than European Americans on most clinical scales but these differences did not occur among psychiatric or substance abusing populations (Greene, 1987, 1991). Differences between Latino Americans and European Americans have generally been found to be less than African American or European American differences (Greene, 1991). The largest difference was that male Latinos scored higher on scale 5 than

male European Americans, (G. Hall et al., 1999). However, all differences were still less than 5 *T*-score points (G. Hall et al., 1999). Given the reviews of ethnicity and the MMPI/MMPI-2 (Greene, 1987, 1991; G. Hall et al., 1999; Schinka et al., 1998), the following conclusions seem warranted:

- Even when ethnic differences have been found between various groups, overall these differences are less than 5 *T*-score points (less than 10% of the variance) and are therefore not clinically meaningful.
- It would be premature to develop new norms for ethnic groups, particularly since moderator variables (SES, age) seem to explain most of the variance in performance.
- It may at times be useful to consider the meanings of ethnic score differences for specific ethnic subgroups. For example, Latino workers compensation cases may be more likely to somatize psychological distress as reflected by greater elevations on 1 (*Hs*), 2 (*D*), and 3 (*Hy*) than European Americans (DuAlba & Scott, 1993). In addition, higher *CYN* (Cynicism) and *ASP* (Antisocial Practices) found among African American as opposed to European American forensic populations are likely to represent clinically meaningful differences (Ben-Porath, Shondrick, & Stafford, 1995).
- Future research should consider within group ethnic differences including degree of identification with his or her ethnic group, language fluency, perceived minority status, and degree to which they feel discriminated.
- More research needs to investigate the relationship between ethnicity and the many supplementary and content scales.

In addition to using the MMPI/MMPI-2 with ethnic groups within the United States, it has also been used in a wide variety of different countries. An important rationale for this is that it is more efficacious to adapt and validate the MMPI/MMPI-2 for a different country than go to the far more extensive effort of developing a whole new test for the culture. Examples of countries where adaptations have occurred include such diverse areas as China, Israel, Pakistan, South Africa, Chili, Mexico, and Japan (see Butcher, 1996). Whenever clinicians work with different cross-national groups, they should consult the specific norms that have been developed for use with these groups, as well as become familiar with any research that may have been carried out with the MMPI on these groups. Useful sources are Butcher's (1996) *International Adaptations of the MMPI-2* and reviews of cross-cultural research by Greene (1987, 1991, pp. 338–354) and G. Hall et al. (1999).

Social Class and Education

The original MMPI was standardized on a group that, by today's standards, would be considered to be poorly educated. There were also additional differences including no or few ethnic groups, relatively lower endorsement of "unusual" item content, and the impact of considerable social changes over the past six decades. The result was that social class, in particular, had to be taken into account when interpreting the original MMPI since it seemed to influence Scales *L, F, K,* and 5 quite extensively. However, the MMPI-2's more representative sample means that, for most educational levels,

education does not need to be taken into account. However, for persons with quite low educational levels (less than 11 years), *K* and 5 (*Mf*) may have lower overall scores so that they need to be interpreted with caution.

The advantages and cautions for using the MMPI, MMPI-2, and MMPI-A indicate that a considerable degree of psychological sophistication is necessary. Both their assets and limitations need to be understood and taken into account. The limitations for the original MMPI are numerous and include moderately adequate reliability, problems related to scale construction, excessive length, offensive items, limited usefulness for normal populations, misleading labels for the scales, inadequacy of the original normative sample, and the necessity of considering demographic variables. Some of these limitations have been corrected by the MMPI-2 and MMPI-A, including an option to decrease the length (by giving only the first 370 items for the MMPI-2 or first 350 items for the MMPI-A), increased appropriateness for normal populations, rewording of 82 of the items, and the use of a larger, more broad-based, modern normative sample. The limitations of the MMPI are also balanced by a number of significant assets, especially the extensive research relating to the meanings of the different scales and the relationships among scales. Extensive strategies are also in place to help refine and expand the meanings of scale elevations by using alternative scales (content, Harris-Lingoes, supplementary). Further assets are the MMPI's familiarity in the field, the development of subgroup norms, and extensive research in specific problem areas. Of central importance is that the MMPI has repeatedly proven itself to have practical value for clinicians, especially because the variables that the scales attempt to measure are meaningful and even essential areas of clinical information. The over 10,000 studies on or using it, combined with its extensive clinical use provide ample evidence of its popularity. The 1989 restandardization should ensure that it not only continues to achieve the status of being an essential psychopathology inventory but will also be a more modern clinical tool.

ADMINISTRATION

The MMPI/MMPI-2 can be administered to persons who are 16 years of age or older with an eighth-grade reading level. As noted, it is possible to administer the MMPI/MMPI-2 to persons between the ages of 16 and 18, but adolescent norms need to be used. However, the preferred option for individuals between ages 14 and 18 is to have them take the MMPI-A. It is often helpful to augment the standard instructions on the MMPI-2 and MMPI-A booklets. In particular, examiners should explain to clients the reason for testing and how the results will be used. It might also be pointed out that the test was designed to determine whether someone has presented himself or herself in an either unrealistically positive or exaggeratedly disturbed manner. Thus, the best strategy is to be as honest and as clear as possible. Finally, it might be clarified that some, or even many, of the questions might seem a bit unusual. They have been developed to assess individuals with a wide range of personality styles and problem presentations. If they don't apply to the person taking the test, this should be indicated with either a true or false response. Including this additional information is likely to result in less anxiety, more accurate responses, and greater rapport. Completion times for all persons taking the test should be noted.

The MMPI-2 and MMPI-A have only one booklet form, although these are available in either softcover or hardcover. Completion of the first 370 items on the MMPI-2 and first 350 items on the MMPI-A allows for the scoring of the basic validity and standard clinical scales. The final 197 MMPI-2 and 128 MMPI-A items are used for scoring different supplementary and content scales. An online computer administration is available through National Computer Systems. For persons who have special difficulties, an individual (Box) form and a tape-recorded form have been developed. The Box form is most appropriate for persons who have difficulties concentrating and/or reading. Each item is presented on a card, which the person is requested to place into one of three different sections to indicate a "true," "false," or "cannot say" response. The tape-recorded form is used for persons who have reading difficulties because of factors such as illiteracy, blindness, or aphasia.

The sometimes prohibitive length of the MMPI has encouraged the development of numerous short forms. However, none is sufficiently reliable or valid to be considered a substitute for the complete administration (Butcher & Hostetler, 1990; Butcher & Williams, 1992; J. Graham, 2000). The only acceptable abbreviated form is to administer all the items necessary for scoring only the basic validity and standard clinical scales (e.g., the first 370 MMPI-2 items or the first 350 MMPI-A items).

Some clinicians allow the client to take the MMPI under unsupervised conditions (such as at home). Butcher and Pope (1989) stress that this is not recommended, for the following reasons:

- The conditions are too dissimilar from those used for the normative samples and any significant change in proceedings might alter the results.
- Clients might consult others to determine which answers to make.
- The clinician cannot be aware of possible conditions that might compromise reliability and validity.
- There is no assurance that the client will actually complete the protocol himself or herself.

Thus, any administration should closely follow the administration procedures used for the normative samples. This means providing clear, consistent instructions, ensuring that the directions are understood, providing adequate supervision, and making sure the setting will enhance concentration by limiting noise and potential interruptions.

INTERPRETATION PROCEDURE

The following eight steps are recommended for interpreting MMPI-2/MMPI-A profiles. These steps should be followed with a knowledge and awareness of the implications of demographic variables such as age, culture, intellectual level, education, social class, and occupation. A summary of the relationship between MMPI profiles and some of the main demographic variables including age, culture, and intellectual level has already been provided. While looking at the overall configuration of the test (Steps 4, 5, and 6), clinicians can elaborate on the meanings of the different scales and the relationships

among scales by consulting the interpretive hypotheses associated with them. These can be found in later sections of this chapter on validity scales, clinical scales, and two-point codes, as well as in sections on supplementary scales and content scales. The discussion of the various scales and codes represents an integration and summary of both primary sources and the following MMPI-2/MMPI-A resources: Archer (1992a), Butcher (1999), Butcher et al. (1989), Butcher et al. (1990), Caldwell (1988), J. Graham (2000), Greene (2000), Greene and Clopton (1994), Keiller and Graham (1993), and Friedman, Lewak, Nichols, and Marks (2000). In particular, the subsections on treatment implications have drawn on the work of Butcher (1990), Greene and Clopton (1994), and Freidman et al. (2000). Occasionally, additional quite recent material and/or relevant reviews/meta-analyses have been cited to either update material related to scale descriptions or highlight important areas of research.

Step 1. Completion Time

The examiner should note the length of time required to complete the test. For a mildly disturbed person who is 16 years or older with an average IQ and eighth-grade education, the total completion time for the MMPI-2 should be approximately 90 minutes. Computer administrations are usually 15 to 30 minutes shorter (60 to 75 minutes in total). The MMPI-A usually takes 60 minutes to complete with computer administrations taking 15 minutes less time (45 minutes in total). If two or more hours are required for the MMPI-2 or 1.5 or more for the MMPI-A, the following interpretive possibilities must be considered:

- Major psychological disturbance, particularly a severe depression or functional psychosis.
- Obsessive indecision.
- Below-average IQ or poor reading ability resulting from an inadequate educational background.
- Cerebral impairment.

If, on the other hand, an examinee finishes in less than an hour, the examiner should suspect an invalid profile, an impulsive personality, or both.

Note any erasures or pencil points on the answer sheet. The presence of a few of these signs may indicate that the person took the test seriously and reduces the likelihood of random marking; a great number of erasures may reflect obsessive-compulsive tendencies.

Step 2. Score and Plot the Profile

Complete the scoring and plot the profile. Specific directions for tabulating the MMPI-2 raw scores and converting them into profiles are provided in Appendix G on page 690. If examiners would like to score and profile the MMPI-2/MMPI-A content scales, Harris-Lingoes and *Si,* or the most frequently used supplementary scales, additional

keys and profile forms may be obtained through National Computer Systems. In addition to the possibility of scoring alternative scales, clinicians should compile further information, including IQ scores, relevant history, demographic variables, and observations derived from Steps 1 and 2.

Score the critical items (see Appendix H on p. 691) and note which ones indicate important trends. It is often helpful at some point to review these items with the client and obtain elaborations. In particular, it is essential to determine whether the person understood what the item was asking. Similarly, it can sometimes be helpful to examine the answer sheet and note which, if any, questions were omitted. A discussion with the client about why he or she chose not to respond might shed additional light on how he or she is functioning psychologically and what areas are creating conflict for him or her.

Step 3. Organize the Scales and Identify the Code Type

The scores can be summarized by simply listing the scores according to the order in which they appear on the profile sheet (*L, F, K,* 1, 2, 3, etc.) with their *T* scores to the right of these scales. For the purposes of communicating scale scores, *T* scores rather than raw scores should be used.

Developing summary codes ("code types") provides a shorthand method of recording MMPI-2/MMPI-A results. Code types can be determined by simply looking at the two highest scale elevations. For example, the two highest scores in a profile might be 8 and 7 resulting in an 87/78 code type. The 87/78 code type can then be looked up in the Two-Point Codes section to obtain various descriptions relating to that code type. Note that Scales 5 (Masculinity-Femininity) and 0 (Social Introversion) are not strictly clinical scales, so they are not used in determining code type. Examiners should keep in mind that only well-defined code types can be safely interpreted (Butcher, 1999; D. Edwards et al., 1993; Greene, 2000; McNulty et al., 1998; Tellegen & Ben-Porath, 1993). A well-defined code type is considered one in which the elevated scales are above 65 and the scales used to determine the code type are 5 or more *T*-score points above the next highest scales. Less well-defined profiles should be interpreted by noting each scale that is elevated and then integrating the meanings derived from the different descriptors.

Step 4. Determine Profile Validity

Assess the validity of the profile by noting the pattern of the validity scales. There are a number of indicators suggesting invalid profiles, which are described in the next section. However, the basic patterns include a defensive style in which pathology is minimized (elevated *L* and/or *K* on the MMPI-2 and MMPI-A), an exaggeration of pathology (elevated *F* and/or *Fb* on the MMPI-2 or *F, F*1, or *F*2 on the MMPI-A), or an inconsistent response pattern (elevated VRIN or TRIN). In addition, clinicians should consider the context of the assessment to determine whether a defensive, fake bad, or inconsistent response style supports what is known about the client. In particular, the examiner should determine the likelihood that the examinee would potentially gain by over- or under-reporting psychopathology.

Step 5. Determine Overall Level of Adjustment

Note the number of scales over 65 and the relative elevation of these scales. The degree to which *F* is elevated can also be an excellent indicator of the extent of pathology (assuming that it is not so high as to indicate an invalid profile). The greater the number and relative elevation of these scales, the more the individual is likely to have difficulties carrying out basic responsibilities and experience social and personal discomfort.

Step 6. Describe Symptoms, Behaviors, and Personality Characteristics

This step represents the core process in interpretation. Mild elevations on individual scales (*T* = 60–65) represent tendencies or trends in the individual's personality. Interpretations should be treated cautiously with the more extreme descriptors being deleted or rephrased to represent milder characteristics. Scores in this range on the MMPI-A are highlighted by shading, thereby designating a marginal or transitional zone between normality and pathology. Elevations above 65 on the MMPI-2 and MMPI-A are more strongly characteristic of the individual and, with progressively greater increases, are more likely to represent core features of personality functioning. However, basing interpretations solely on specific *T* score elevations may be misleading because a client's demographic characteristics often exert a strong influence. For example, persons with lower educational backgrounds usually score lower on *K* and *Mf;* therefore, interpretations need to take this into account. Furthermore, different authors use different criteria for determining high and low scores. Some authors have used *T* score ranges (e.g., *T* = 70–80); others have defined elevated scores as the upper quartile; and still others have defined a high score as the highest in a profile regardless of other *T* score elevations. This issue is further complicated because two persons with the same elevation on a scale but with quite different personal histories (e.g., psychiatric vs. adequate level of functioning) will have different interpretations that are appropriate for them. As a result, the descriptors in the following sections on interpretation do not designate specific *T* score elevations. Instead, more general descriptions associated with high and low scores have been provided. Clinicians will need to interpret the accuracy of these potential meanings by taking into consideration not merely the elevations, but other relevant variables as well. In addition, each of the descriptions are modal. They should be considered as possible interpretations that will not necessarily apply to all persons having a particular score. They are merely hypotheses in need of further verification. This is highlighted by the finding that somewhere in the range of 40% of computer-generated descriptors do not apply to the person being assessed (Butcher et al., 2000).

Whereas *T* scores are not provided for most scale interpretations, they have been included in the subsection on validity scales. Validity *T* and sometimes raw scores are included because there is extensive research on optimal cutoff scores.

During the interpretive process, do not merely note the meanings of the individual scales but also examine the overall pattern or configuration of the test and note the relative peaks and valleys. Typical configurations, for example, might include the "conversion V," reflecting a possible conversion disorder or elevated Scales 4 and 9, which

reflect a high likelihood of acting-out behavior. Note especially any scales greater than 65 or less than 40 as being particularly important for the overall interpretation. The meaning of two-point code configurations can be determined by consulting the corresponding section in this chapter (Two-Point Codes). When working to understand the meaning of a profile with two or more elevated clinical scales, it is recommended that clinicians read the descriptors for the individual scales, as well as relevant two-point code descriptions. It is also recommended that, when reading about elevations on single scales, clinicians should read the meanings of high and low elevations, as well as the more general information on the relevant scale. Further elaboration on the meaning of the scale elevations and code types can be obtained by scoring and interpreting the content scales, Harris-Lingoes and *Si* subscales, supplementary scales, and/or the critical items; these scales are discussed later in this chapter. J. Graham (2000) recommends that, when possible, descriptions related to the following areas should be developed: test-taking attitude, adjustment level, characteristic traits/behaviors, dynamics/etiology, diagnostic impressions, and treatment implications. When interpretive information is available, clinicians can examine an individual's profile in combination with the requirements of the referral questions to determine relevant descriptions for each of these areas.

Clearly Defined Profiles

As noted previously, a clearly defined code type is indicated by both a high elevation and either single scales, which are elevated with no other "competing" scale elevations (so-called spike profiles), or clear code types in which the elevated scales in the code types similarly do not have competing scales that are close to the degree of elevations of the scales in the code. Well-defined elevations indicate greater validity of the relevant descriptors (McNulty et al., 1998). In addition, they are more likely to be stable over time (high test-retest reliability).

Poorly Defined Profiles

If the elevation is not particularly high (generally $T = 60–65$), the interpretations need to be modified by either toning down the descriptors to a more normal level, or deleting the more extreme descriptors. Often the content, Harris-Lingoes, and supplementary scales can be useful in understanding the meaning of elevations in the $T = 60–64$ range. If the profile is poorly defined because there are additional scales that "compete" with the scales in the code type (e.g., 27/72 code type but with Scales 1 and 8 also elevated nearly as high as Scales 2 and 7), several strategies need to be used. The safest and most conservative strategy is to consider descriptors that occur in common among all the different elevated scales as the most valid (e.g., anxiety is likely to be a common descriptor for elevations on Scales 1, 2, 7, and 8; this is strengthened if 7 is the most highly elevated scale). In addition, examiners need to make an effort to understand and integrate the interpretations given under each of the individual scale descriptions. Furthermore, the meanings of alternative code type combinations need to be considered and integrated (e.g., if Scales 2, 7, 1, and 8 are all elevated, the following code type descriptors need to be considered: 27/72, 18/81, 87/78, 12/21, 17/71, and 28/82). Finally, with poorly defined elevations, it becomes increasingly important to use the content, Harris-Lingoes, critical items, and supplementary scales to more fully understand and refine the meanings of the clinical scale elevations.

Use of Content Scales

The content scales can be used to supplement, extend, confirm, and refine interpretations derived from the basic validity and standard clinical scales. Furthermore, some of the content scales (e.g., TPA/Type A, WRK/Work Interference) provide additional information not included in the clinical scales. The adult content scales are divided into the clusters of internal symptoms, external aggressive tendencies, negative self-view, and general problem areas. Similarly, the adolescent content scales are divided into scales reflecting interpersonal functioning, treatment recommendations, and academic difficulties (see Content Scales section).

Harris-Lingoes and Si *Subscales*

To understand which personality and clinical variables of a person might have been responsible for elevating the clinical scales, clinicians might wish to selectively use the rationally devised Harris-Lingoes and Social Introversion subscales. These scales (or subscales) organize clusters of content-related items so that the different dimensions of the scales can be more clearly differentiated. For example, it might be found that an elevation on Scale 4 (Psychopathic Deviate) resulted primarily from family discord. In contrast, criminal acting out might be suggested by subscale elevations on authority conflict and social imperturbability. This would then have implications for both interpretations and case management (see Harris-Lingoes Scales section).

Critical Items

Clinicians may also wish to evaluate the meanings of content related to specific items the client has endorsed by investigating critical items (see Critical Items section).

Supplementary Scales

The empirically derived supplementary scales can also be used to both refine the meanings of the clinical scales, and add information not included in the clinical scales (see Supplementary Scales section).

Low Scale Scores

For the most part, low scale scores (below *T* score of 35 or 40) are likely to represent strengths and these may serve to modify any high scale elevations (see sections on low scores for each scale).

Specific Interpretive Guidelines Organized around Symptom Domains

Suppression (constriction). Scales 5 (*Mf*) and 0 (*Si*) are sometimes referred to as suppressor scales because, if either or both are elevated, they tend to suppress or "soften" the expression of characteristics suggested by other elevated scores.

Acting Out (impulsivity). In contrast to Scales 5 (*Mf*) and 0 (*Si*), Scales 4 (*Pd*), and 9 (*Ma*) are sometimes referred to as "releaser" or "excitatory scales"; and if one or both are elevated, the person is likely to act out difficulties. This hypothesis is further strengthened if 0 (*Si*) is also quite low.

Internalizing Coping Style. Similar to the preceding two guidelines are indicators of internalizing versus externalizing coping styles. If the combined scores for Scales

4 (*Pd*), 6 (*Pa*), and 9 (*Ma*) are lower than the combined scores for 2 (*D*), 7 (*Pt*), and 0 (*Si*), the individual can be considered to have an internalizing coping style.

Externalizing Coping Style. In contrast to the preceding, an individual who has combined scores on 4 (*Pd*), 6 (*Pa*), and 9 (*Ma*) that are greater than his or her combined scores on 2 (*D*), 7 (*Pt*), and 0 (*Si*) can be considered to have an externalizing coping style.

Overcontrol (repression). Rigid overcontrol of impulses, particularly hostility, is suggested by elevations on 3 (*Hy*) and the *O-H* (Overcontrolled Hostility) supplementary scale.

Anger (loss of control). Angry loss of control is suggested by elevations on the ANG (Anger) content scale.

Subjective Distress. A general check on the degree of subjective stress a person is encountering can be determined by noting the degree to which scales 2 (*D*) and 7 (*Pt*) are elevated.

Anxiety. Elevations on Scale 7 (*Pt*), especially if 7 (*Pt*) is greater than 8 (*Sc*), suggest anxiety.

Depression. A high score on 2 (*D*) combined with a low score on 9 (*Ma*) is particularly indicative of depression.

Mania. A high score on 9 (*Ma*) combined with a low score on 2 (*D*) suggests mania.

Psychosis. A high score on 8 (*Sc*), especially if 8 (*Sc*) is 10 points or more higher than 7 (*Pt*), suggests psychosis.

Confusion and Disorientation. Elevations above $T = 80$ on *F,* 8 (*Sc*), and 7 (*Pt*) suggest a confused, disoriented state. Confusion can also be suggested if the mean for all 8 clinical scales (this excludes Scales 5 and 0 as these are not strictly clinical scales) is greater than $T = 70$.

Suspicion Mistrust. If 6 (*Pa*) is moderate to highly elevated and, especially if 6 is the highest scale, suspicion and mistrust is strongly indicated.

Introversion. Introversion is indicated by elevations on the 0 (*Si*) scale.

Obsessiveness. Obsessiveness is indicated by elevations on 7 (*Pt;* especially when this is the highest point) and elevations on the *OBS*/Obsessiveness content scale.

Cynicism. Cynicism is indicated by elevations on the CYN (Cynicism) content scale.

Drug or Alcohol Problems. Elevations on Scales 4 (*Pd*), 2 (*D*), and 7 (*Pt*) are consistent with (although not diagnostic of) drug- and alcohol-related problems. Lifestyle and personality patterns consistent with, and suggesting proneness to drug and alcohol patterns, are indicated by elevations on MAC-R and the Alcohol Potential Scale (APS). Clear awareness of and open discussion of alcohol and/or drug problems are indicated by elevations on the Alcohol Acknowledgment Scale (AAS).

Quality and Style of Interpersonal Relations. Scales that are most useful for understanding the patterns of interpersonal relations include the following:

- 0 (*Si;* level of sociability, shyness, social avoidance, alienation).
- Social Discomfort Scale (SOD; social discomfort).
- 1 (*Hs;* complaining, critical, demanding, indirect expression of hostility, passive, preoccupied with self).

- 4 (*Pd;* good first impressions but use others for their own needs, outgoing, talkative, energetic but also shallow and superficial, and impulsive).
- 6 (*Pa;* moralistic, suspicious, hypersensitive, resentful, guarded).
- 8 (*Sc;* isolated from social environment, seclusive, withdrawn, inaccessible, feels misunderstood).
- Marital Distress Scale (MDS; presence of marital distress).
- Dominance (*Do;* assertive, dominant, takes the initiative, confident).

The preceding topic areas and interpretive strategies are intended to be basic, rule-of-thumb approaches to help guide hypothesis generation around specific areas. There are certainly other relevant areas, but the ones listed can generally be considered the most important. While these guidelines will serve to alert clinicians to specific areas, they will still need to investigate these areas in far more depth by consulting relevant scale descriptors and patterns between scales. Clinicians may also wish to consult one of the MMPI-2/MMPI-A resources listed in the recommended readings to further extend and expand on the meanings of different profiles.

Step 7. Provide Diagnostic Impressions

Although the original MMPI as well as the MMPI-2/MMPI-A have not been successful in leading directly to diagnosis, they can often contribute considerable information relevant to diagnostic formulations. In the section on code types, possible *DSM-IV* diagnoses consistent with each code type have been included. Clinicians should consider these, along with additional available information, to help make an accurate diagnosis. In some contexts and for some types of referral questions, this will be relevant; but for other contexts and referral questions, formal diagnosis will be neither required nor appropriate (e.g., employment screening). A further review of the considerations and guidelines described in Step 6 might be useful in extracting relevant information for diagnosis.

Step 8. Elaborate on Treatment Implications and Recommendations

Often, one of the most valuable services a practitioner can provide is to predict the client's likelihood of benefiting from interventions. This typically means elaborating on the person's strengths and weaknesses, level of defensiveness, ability to form a treatment relationship, predicted response to psychotherapy (note especially *Es* [ego strength] and TRT scales), antisocial tendencies, and level of insight. Much of this information is summarized at the ends of the subsections on scale elevations and code types. If doing extensive work with specific types of clients, clinicians might need to expand on the knowledge relating to types and outcome of treatments by referring to the extensive research base that is available (e.g., chronic pain, substance abuse, outcomes related to specific code types). Butcher's (1990) *MMPI-2 in Psychological Treatment* can be particularly helpful in this regard. Treatment responsiveness might be further extended into providing suggestions for tailoring specific interventions for client profiles and types of problems. Reviewing the areas, considerations, and guidelines

described in Step 6 might be useful in extracting information relevant to treatment planning. A further useful resource in this process is Maruish's (1999) *The Use of Psychological Testing for Treatment Planning and Outcome Assessment*. Lewak, Marks, and Nelson (1990) not only provide implications for treatment but also outline a step-by-step procedure for translating MMPI-2 results into clear, relevant feedback for the client. These steps include specific issues for the client background and early life experiences and self-help suggestions. A listing of representative feedback statements derived from Lewak et al. (1990) is provided in Appendix H (pp. 691–693). These can be used to provide feedback in a user friendly manner which will be likely to increase rapport and optimize client growth. In addition, a manual by Finn (1996) for using the MMPI-2 as a therapeutic intervention is available.

COMPUTERIZED INTERPRETATION

Computerized interpretation systems are an important and frequently used adjunct to MMPI interpretation. The number of such services has grown considerably since 1965 when the first system was developed by the Mayo Clinic. Major providers are National Computer Systems, Psychological Assessment Resources, Roche Psychiatric Service Institute (RPSI), Clinical Psychological Services, Inc. (using the Caldwell Report), Western Psychological Services, and Behaviordyne Psychodiagnostic Laboratory Service. A description and evaluation of many of these services are included in the *Mental Measurements Yearbook* (the most recent/13th edition was edited by Impara & Plake, 1998). Early lists and descriptions of software packages available for personal computers can be found in Krug's (1993) *Psychware Sourcebook* (4th ed.) or Stoloff and Couch's (1992) *Computer Use in Psychology: A Directory of Software* (3rd ed.). However, because of rapid changes in computer offerings, the best sources will be the most recent listings found in catalogues of the major distributors (see Appendix A on p. 673).

Caution in the use of different computer-based interpretive systems is important because the interpretive services and software packages are highly varied in terms of quality, and most of them have untested or only partially tested validity. Many do not specify the extent to which they were developed using empirical guidelines versus clinical intuition. Each computerized system has a somewhat different approach. Some provide screening, descriptive summaries, and cautions related to treatment, whereas others provide extensive elaborations on interpretations or may provide optional interpretive printouts for the clients themselves. Even the best programs will produce a combination of accurate as well as inaccurate interpretations (Butcher et al., 2000).

The rationale behind computerized systems is that they are efficient and can accumulate and integrate large amounts of information derived from the vast literature on the MMPI, which even experienced clinicians cannot be expected to recall. However, questions have been raised regarding misuse (Groth-Marnat, 1985; Groth-Marnat & Schumaker, 1989). In particular, computerized services are limited to standard interpretations and are not capable of integrating the unique variables usually encountered in dealing with clinical cases. This is a significant factor, which untrained personnel may be more likely to either overlook or inadequately evaluate. In response to these issues, the American Psychological Association developed a set of guidelines to ensure

the proper use of computerized interpretations (American Psychological Association, 1986, 1994). It should be stressed that, although computerized systems can offer information from a wide variety of accumulated data, their interpretations are still not end products. Like all test data, they need to be placed in the context of the client's overall background and current situation, and integrated within the framework of additional test data (see McMinn, Ellens, et al., 1999).

VALIDITY SCALES

The MMPI was one of the first tests to develop scales to detect whether respondents were answering in such a manner as to invalidate the overall results. This tradition has continued and been expanded into the newer MMPI-2 and MMPI-A. Meta-analyses of studies on the various validity scales generally indicate that they are able to effectively detect faking. Probably the most effective strategy is the *F* scale's ability to detect overreporting of pathology (R. Baer, Kroll, Rinaldo, & Ballenger, 1999; Bagby, Buis, & Nicholson, 1995; Iverson, Franzen, & Hammond, 1995; G. Meyer & Archer, 2001). The *K* scale, while still useful, is somewhat less effective in detecting underreporting (R. Baer, Wetter, & Berry, 1992; Putzke, Williams, Daniel, & Boll, 1999). However, adding supplementary validity scales (Social Desirability scale, Superlative scale) to *L* and *K* can serve to increase the detection of underreporting (Bagby, Rogers, Nicholson, et al., 1997). Despite the consensus related to the accuracy of detection, a concern is that a wide range of cutoff scores are recommended depending on the group being assessed (Bagby et al., 1994, 1995; L. Stein, Graham, & Williams, 1995). For example, optimal cutoff scores for normals faking bad are lower than psychiatric patients faking bad (Berry, Baer, & Harris, 1991; J. Graham et al., 1991). An unresolved issue is whether normals who are motivated to fake bad as well as given information on how to fake (e.g., symptom patterns of individuals with posttraumatic stress disorder, paranoid schizophrenia, schizophrenia) can avoid detection. Some research indicates that, even with motivation and a clear strategy, they still cannot avoid detection (Wetter, Baer, Berry, Robinson, & Sumpter, 1993), whereas other research suggests that strategic (informed) fakers can consistently produce profiles that are indistinguishable from true patients (R. Rogers, Bagby, & Chakraborty, 1993; Wetter & Deitsch, 1996). Attempts to fake bad might be particularly likely to succeed if subjects are given information on the design and intent of the validity scales (Lamb, Berry, Wetter, & Baer, 1994) and are familiar with the type of disorder they are faking (Bagby, Rogers, Buis, et al., 1997).

It should be noted that the MMPI-2 provides the option of profile sheets that can include either *K* corrections or sheets that omit this procedure. The MMPI-A does not include the *K* correction on its profile sheets because, in some contexts, particularly those for adolescents, the *K* correction is not appropriate (Colby, 1989).

The ? "Scale" (Cannot Say; *Cs*)

The ? scale (abbreviated by either ? or *Cs*) is not actually a formal scale but merely represents the number of items left unanswered on the profile sheet. The MMPI-2 does not

even include a column for profiling a ? (*Cs*) scale, but merely provides a section to include the total number of unanswered questions. The usefulness of noting the total number of unanswered questions is to provide one of several indices of a protocol's validity. If 30 or more items are left unanswered, the protocol is most likely invalid and no further interpretations should be attempted. This is simply because an insufficient number of items have been responded to, which means less information is available for scoring the scales. Thus, less confidence can be placed in the results. To minimize the number of "cannot say" responses, the client should be encouraged to answer all questions.

A high number of unanswered questions can occur for a variety of reasons. It might indicate difficulties with reading, psychomotor retardation, indecision, confusion, or extreme defensiveness. These difficulties might be consistent with severe depression, obsessional states, extreme intellectualization, or unusual interpretations of the items. Defensiveness might stem from legalistic overcautiousness or a paranoid condition. High *Cs* might also occur from the perception that the unanswered items are irrelevant.

VRIN (Variable Response Inconsistency Scale)

The VRIN scale is a new MMPI-2 and MMPI-A validity scale designed to complement the existing validity scales. It comprises pairs of selected questions that would be expected to be answered in a consistent manner if the person is approaching the testing in a valid manner. Each pair of items is either similar or opposite in content. It would be expected that similar items would be answered in the same direction. If a person answers in the opposite direction, then it indicates an inconsistent response and is, therefore, scored as one raw score on the VRIN scale. Pairs of items with opposite contents would be expected to be answered in opposite directions. If, instead, these pairs are answered in the same direction, this would represent inconsistent responding, which would also be scored as one raw score point on the VRIN scale.

High VRIN (MMPI-2 $T = 80$; MMPI-A $T = 61$)

A high number of inconsistent responses suggests indiscriminate responding. Thus, the profile would be considered invalid and should not be interpreted. *T* scores 5 to 10 points below this (MMPI-2 70-79; MMPI-A 70-74) also raise the possibility that the profile is invalid. If VRIN is high along with a high *F,* this further suggests that the person has answered in a random manner. In contrast, a low or moderate VRIN accompanied by high *F* suggests that the person was either severely disturbed, or was intentionally attempting to exaggerate symptoms.

TRIN (True Response Inconsistency Scale)

The MMPI-2 and MMPI-A TRIN scales are like the VRIN scale in comprising pairs of items. However, only pairs with opposite contents are included. This means there would be two ways for a person to obtain a response that would be scored on the VRIN scale. A "True" response to both items would indicate inconsistency and would, therefore, be scored as plus one raw score point. A "False" response to both pairs would also indicate

inconsistency but would be scored as minus one point (negative scores are avoided by adding a constant).

Very High Scores (MMPI-2 $T = 80$; MMPI-A $T = 75$)

Scores in this range indicate that the person is indiscriminately answering "True" to the items (acquiescence or yea-saying). However, *T* scores 5 to 10 points (MMPI-2 70–79; MMPI-A 70–74) below the indicated cutoffs can still raise the possibility of an invalid profile. Thus, an elevated TRIN is interpreted in much the same way as elevated VRIN profiles. Correspondingly, a high *F,* accompanied by a high TRIN, suggests indiscriminate responses, whereas a high *F* and a low to moderate TRIN suggest either excessive pathology or an exaggeration of symptoms.

The *F* Scale (Infrequency)

The *F* (Infrequency) scale measures the extent to which a person answers in an atypical and deviant manner. The MMPI and MMPI-2 *F* scale items were selected based on their endorsement by less than 10% of the population. Thus, from a statistical definition, they reflect nonconventional thinking. For example, a response is scored if the client answers "True" to item 49, "It would be better if almost all laws were thrown away" or "False" to 64, "I like to visit places where I have never been before." However, the items do not cohere around any particular trait or syndrome. This indicates that a client who scores high is answering in a scorable direction to a wide variety of unusual characteristics. As might be expected, high scores on *F* are typically accompanied by high scores on many of the clinical scales. High scores can often be used as a general indicator of pathology. In particular, high scores can reflect unusual feelings caused by some specific life circumstance to which the person is reacting. This might include grieving, job loss, or divorce. A person scoring high may also be "faking bad," which could serve to invalidate the protocol. No exact cutoff score is available to determine whether a profile is invalid or is accurately reflecting pathology. Even *T* scores from 70 to 90 do not necessarily reflect an invalid profile, particularly among prison or inpatient populations. In general, moderate elevations represent an openness to unusual experiences and possible psychopathology, but it is not until more extreme elevations that an invalid profile is suspected. Further information can be obtained by consulting the *F* back scale (see sections on *F* back or *Fb*).

The 66-item MMPI-A *F* scale was constructed similar to the MMPI-2 *F* scale. However, because adolescents are more likely to endorse unusual experiences, a more liberal criterion of 20% endorsement was used for inclusion. The MMPI-A *F* scale was further divided into *F*1 scales to assess validity for the first portion of the booklet (clinical scales) and *F*2 to assess the last portion of the book (supplementary and content scales; see sections on F-K index, *F*1 and *F*2).

High Scores on F

MMPI-2 scores approximating 100 or greater suggest an invalid profile. This might have been caused by clerical errors in scoring, random responding, false claims by the client regarding symptoms, or distortions caused by a respondent's confused and

delusional thinking. In rare cases where psychiatric scores of 100 do reflect a valid assessment, it indicates hallucinations, delusions of reference, poor judgment, and/or extreme withdrawal. Because of the variations in scores between normals and psychiatric patients, different recommendations for indicating an invalid profile have been provided for each of these groups. For normal adult populations faking bad, an MMPI-2 raw cutoff score of 18 (male $T = 95$; female $T = 99$) is recommended, whereas for psychiatric patients faking bad, a higher raw cutoff score of 29 ($T = 120$) is recommended for males and 27 ($T = 119$) for females (J. Graham et al., 1991). Optimal cutoff scores for the MMPI-A are $T = 79+$ (R. Baer et al., 1999).

In evaluating the possibility of an invalid profile, MMPI-2 *T* scores of 80 to 99 suggest malingering, an exaggeration of difficulties, resistance to testing, or significant levels of pathology. In those cases where this range accurately reflects pathology, clients should have corresponding histories and behavioral observations indicating they are disoriented, restless, moody, and dissatisfied. Scores of 70 to 80 suggest the client has unconventional and unusual thoughts and may be rebellious, antisocial, and/or having difficulties in establishing a clear identity. Scores from 70 to 90 might represent a "cry for help," in which persons are being quite open regarding their difficulties in an attempt to indicate they need assistance. If a client scores from 65 to 75, but does not seem to be pathological, he or she might be curious, complex, psychologically sophisticated, opinionated, unstable, and/or moody.

Low Scores on F

Low scores on *F* indicate that clients perceive the world as most other people do. However, if their history suggests pathology, they might be denying difficulties ("faking good"). This distinction might be made by noting the relative elevation on *K* and *L*.

Fb (*F* back) Scale (MMPI-2); *F*1 and *F*2 (MMPI-A)

The 40-item MMPI-2 *Fb* was developed in conjunction with the restandardization of the MMPI. It was designed to identify a "fake bad" mode of responding for the last 197 items. This might be important because the traditional *F* scale was derived only from items taken from what are now the first 370 questions on the MMPI-2. Without the *Fb* scale, no check on the validity of the later questions would be available. It might be possible for a person to answer the earlier items accurately and later change to an invalid mode of responding. This is important for the supplementary and content scales because many of them are derived either partially or fully from the last 197 questions. The *Fb* scale was developed in the same manner as the earlier *F* scale in that items with low endorsement frequency (less than 10% of nonpatient adults) were included. Thus, a high score suggests the person was answering the items in an unusual mode. As with the *F* scale, this could indicate either generalized pathology or that the person was attempting to exaggerate his or her level of symptomatology.

Somewhat similar to the MMPI-2, the MMPI-A includes a 66-item *F* scale that is divided into *F*1 and *F*2 subscales. The *F*1 scale is composed of 33 items, all of which appear on the first half (initial 236 items) of the MMPI-A booklet, and relates to the standard clinical scales. In contrast, the 33-item *F*2 scale is composed of items on the last half of the booklet (final 114 items) and relates to the supplementary and clinical

scales. The *F*1 and *F*2 scales can be interpreted in much the same way as for *F* and *Fb* on the MMPI-2. The optimal cutoff score for *Fb* (and *F*1 and *F*2) is not known. As a result, it is recommended that the same guidelines and *T* scores used to determine an invalid profile for *F* also be used for *Fb* (and *F*1 and *F*2). However, the *Fb* scale has not been found to be as effective a predictor of malingering as the *F* scale (Iverson et al., 1995).

The *Fp* (Infrequency-Psychopathology) Scale

Because the *F* scale is typically elevated among psychiatric patients, it is often difficult to differentiate between persons with true psychopathology and those who have some psychopathology but are nonetheless faking bad. This is particularly true if the psychopathology is quite severe. The history of the person (e.g., degree of preexisting psychopathology) and context of the referral (e.g., possible gain for faking bad) can often be quite useful in making this distinction. To further assist with this differentiation, Arbisi and Ben-Porath (1995) developed a set of 27 items that were infrequently answered even by psychiatric inpatients. (In contrast, the *F* scale was developed from infrequently answered questions by the normative sample.) This means that high scores on *Fp* ($T > 106$ for men and $T > 113$ for women) can potentially identify persons who are faking bad even if they are psychiatric patients (Arbisi & Ben-Porath, 1998; R. Rogers, Sewell, & Ustad, 1995).

The *L* (Lie) Scale

The *L* or *lie* scale consists of 15 items that indicate the extent to which a client is attempting to describe himself or herself in an unrealistically positive manner. Thus, high scorers describe themselves in an overly perfectionistic and idealized manner. The items consist of descriptions of relatively minor flaws to which most people are willing to admit. Thus, a person scoring high on the *L* Scale might answer "False" to item 102, "I get angry sometimes." However, persons from uneducated, lower socioeconomic backgrounds are likely to score somewhat lower on *L* than those from higher socioeconomic backgrounds with more education. On the original MMPI, the *L* score is figured by counting the total number of false responses to items 15, 45, 75, 135, 165, 195, 225, 255, 285, 30, 60, 90, 120, and 150. The MMPI-2 and MMPI-A provide separate scoring keys.

High Scores on L

Evaluating whether an *L* scale is elevated requires that the person's demographic characteristics first be considered. A raw score of 4 or 5 would be a moderate score for lower-class persons or persons from the middle class who are laborers. In contrast, a raw score of 4 or 5 would be considered high for college-educated persons unless it can be explained based on their occupations (e.g., clergy). A high score on the MMPI-A is considered to be $T = 66+$. If the client's score is considered high, it may indicate the person is describing himself or herself in an overly favorable light. This may result from conscious deception or, alternatively, from an unrealistic view of himself or herself. Such clients may be inflexible, unoriginal, and unaware of the impressions they make on others, and perceive their world in a rigid, self-centered manner. As a result of their rigidity, they may have a low tolerance to stress. Because they will deny any

flaws in themselves, their insight will be poor. This is likely to make them poor candidates for psychotherapy. Extremely high scores would suggest that they are ruminative, extremely rigid, and will experience difficulties in relationships. This may be consistent with many paranoids who place considerable emphasis on denying their personal flaws and instead project them onto others. Extremely high scores might also be the result of conscious deception by antisocial personalities.

Low Scores on L

Low scores suggest that clients were frank and open regarding their responses to the items. They are able to admit minor faults in themselves and may also be articulate, relaxed, socially ascendant, and self-reliant. Low scores might also indicate clients who are somewhat sarcastic and cynical.

The *K* (Correction) Scale

The *K* scale was designed to detect clients who are describing themselves in overly positive terms. It, therefore, has a similarity with the *L* scale. The *K* scale, however, is more subtle and effective. Whereas only naive, moralistic, and unsophisticated individuals would score high on *L,* more intelligent and psychologically sophisticated persons might have somewhat high *K* scores and yet be unlikely to have any significant elevation on *L*. Persons from low socioeconomic and educational backgrounds score somewhat lower on *K;* therefore, this might need to be taken into account during interpretation for these groups.

Moderate scorers often have good ego strength, effective emotional defenses, good contact with reality, and excellent coping skills. Typically, they are concerned with, and often skilled in, making socially acceptable responses. As might be expected, *K* scores are inversely related to scores on Scales 8, 7, and 0. Elevations on *K* can also represent ego defensiveness and guardedness. This might occur with persons who avoid revealing themselves because of their personality style or because something might be gained by conveying a highly favorable impression (e.g., employment). There is no clear cutoff for differentiating between positive ego strength (adjustment), ego defensiveness, or faking good. A general guideline is that the more ego-defensive the person is, the more likely it is that some of the clinical scales might also be elevated. Helpful information can also be obtained through relevant history and the context of the testing (legal proceedings, employment evaluation, etc.).

Because a defensive test-taking approach is likely to suppress the clinical scales, a *K* correction is added to five of the MMPI-2 clinical scales (1/*Hs,* 4/*Pd,* 7/*Pt,* 8/*Sc,* 9/*Ma*) to compensate for this defensiveness. This correction is obtained by taking a designated fraction of *K* and adding it to the relevant scale (see directions in Appendix G on p. 690). However, the basis of the *K* correction has been called into question. It has been omitted from the MMPI-A, and the MMPI-2 contains separate scoring sheets with and without the *K* correction so that examiners can decide whether they wish to use it.

High Scores on K

Scores that are much higher than would be expected (generally above $T = 65$ or 70) given the person's history suggest that clients are attempting to describe themselves in an overly favorable light or deny their difficulties, or that they answered false to all

items (nay-saying). If the profile is considered valid, high scores indicate such persons are presenting an image of being in control and functioning effectively, but they will overlook any faults they might have. They will have poor insight and resist psychological evaluation. Because they will resist being in a patient role, their ability to benefit from psychotherapy may be limited. They will be intolerant of nonconformity in others and perceive them as weak. Their typical defense will be denial, and, because of their poor insight, they will be unaware of the impression they make on others. They might also be shy, inhibited, and have a low level of social interaction.

Moderate Scores on K

Moderate scores ($T = 56–64$) suggest moderate levels of defensiveness, as well as a number of potential positive qualities. These clients may be independent, self-reliant, express an appropriate level of self-disclosure, and have good ego strength. Their verbal ability and social skills might also be good. Although they might admit to some "socially acceptable" difficulties, they might minimize other important conflicts. They would be unlikely to seek help. Moderate scores in adolescents contraindicate acting out.

Low Scores on K

Low scores suggest a fake bad profile in which the person exaggerates his or her pathology. It might also suggest a protocol in which all the responses have been marked true. In an otherwise valid profile, the client might be disoriented and confused, extremely self-critical, cynical, skeptical, dissatisfied, and have inadequate defenses. He or she would be likely to have a poor self-concept with a low level of insight. Low scores among adolescents are not uncommon and may reflect a greater level of openness and sensitivity to their problems. It might be consistent with their undergoing a critical self-assessment related to establishing a clear sense of identity. Low scores are also quite common among clients from lower socioeconomic levels and, as such, would not be as likely to suggest difficulties with adjustment.

The F-K Index (Dissimulation Index)

The difference between scores on *F* and *K* has been used to provide an index of the likelihood that a person is producing an invalid profile. This index can be determined by subtracting the raw score (not *T* scores) on *K* from the raw score on *F*. However, most research has indicated that it is not as effective as using the *F* scale to detect fake bad profiles (R. Baer et al., 1999; Bagby et al., 1994, 1995). In addition, not enough is known about the MMPI-2 and MMPI-A F-K index to use it in detecting fake good profiles. As a result, it is not recommended that this index be used.

The *S* (Superlative) Scale

Because the *K* and *L* scales have been found to be only moderately effective in differentiating persons who fake good, the *S* scale was developed in the hopes that it might more accurately identify those persons attempting to appear overly virtuous (Butcher & Han, 1995). The 50 items of the scale were developed by noting the differences in item endorsement between persons in an employment situation who were likely to be presenting themselves in an extremely favorable light (i.e., airline pilots applying for a

job) and the responses of the normative sample. The resulting 50 items relate to contentment with life, serenity, affirming human goodness, denial of irritability/anger, patience, and denial of moral flaws. Thus, persons endorsing a high number of these items are presenting themselves as getting along very easily with others, being free from psychological problems, and having a strong belief in human goodness.

The scale does seem to be effective in discriminating nonpatients who were requested to present themselves in an extremely favorable light (pretending they were applying for a highly desired job) from those who were requested to respond in an honest manner (R. Baer, Wetter, Nichols, et al., 1995). It may also add to the validity of *K* and *L* in identifying persons who are faking good. R. Baer, Wetter, and Berry (1995) reported that a raw cutting score of 29 ($T = 55$) was able to correctly identify 91% of a group of students who were asked to take the MMPI-2 honestly and 92% of those who were requested to respond to the items in an extremely favorable (fake good) manner. However, they also reported that it was not as effective as the *L* scale in distinguishing psychiatric patients who were faking good from patients who were responding honestly (R. Baer, Wetter, & Berry, 1995). In summary, the *S* scale shows some promise in increasing the detection rate of persons faking good beyond merely using *K* and *L* among normal populations (but not among psychiatric patients). At this point, it should be considered an experimental scale and any decisions made on it should be done with considerable caution.

CLINICAL SCALES

Scale 1. Hypochondriasis (*Hs*)

Scale 1 was originally designed to distinguish hypochondriacs from other types of psychiatric patients. Although it can suggest a diagnosis of hypochondriasis, it is most useful as a scale to indicate a variety of personality characteristics that are often consistent with, but not necessarily diagnostic of, hypochondriasis. High scorers show not only a high concern with illness and disease, but also are likely to be egocentric, immature, pessimistic, sour, whiny, and passive-aggressive. They rarely act out directly but, rather, express hostility indirectly and are likely to be critical of others. Their complaints are usually related to a wide variety of physical difficulties. An important purpose of these complaints is to manipulate and control others. Low scores suggest an absence of these complaints.

Scale 1 may also be elevated along with Scales 2, 3, and 7. This would reflect corresponding levels of depression, denial, conversions, or anxiety states. However, persons who score high on Scale 1 typically experience little overt anxiety. A conversion V occurs when there are elevations on Scales 1 and 3 along with a significant lowering (10 or more points) on 2. This profile suggests that the person converts psychological conflicts into bodily complaints (see 13/31 code type).

High Scores on Scale 1

Persons with high scores on Scale 1 are described as stubborn, pessimistic, narcissistic, and egocentric. Symptom complaints typically include epigastric complaints, pain, fatigue, and headaches. While they will have a reduced level of efficiency, they are

rarely completely incapacitated. Others might perceive them as dull, unenthusiastic, ineffective, and unambitious. They will use their symptom-related complaints to manipulate others and, as a result, will make others around them miserable. Their symptom-related complaints are vague and diffuse and will often shift to various locations on their bodies. They often overuse the medical system and their histories might reveal numerous visits to a wide variety of practitioners. However, they refuse to believe assurances that their difficulties have no organic basis. Each time they visit a physician, they will recite a long series of symptom complaints (sometimes referred to as an "organ recital"). Their symptoms are usually not reactions to situational stress but more of long-standing duration. Persons with moderate scores may have a true organic basis for their difficulties. However, even moderately high scorers will be likely to exaggerate their physical difficulties. If Scale 7 is also elevated, this may indicate a better prognosis for psychotherapy because clients' level of anxiety is high enough to motivate them to change. Extremely high scores might suggest that the person has a wide variety of symptom-related complaints and will probably be extremely constricted. This might be consistent with a person with psychotic-like features (schizoid, schizoaffective, schizophrenic, psychotic depression) who is having bodily delusions (check elevations on Scales 6, 7, 8, and 9).

Treatment Implications Individuals with elevations on Scale 1 have often rejected and criticized the "help" that has been offered to them. They would be likely to reset any suggestion that their difficulties are even partly psychologically based. Since their level of insight is quite poor, they typically do not make good candidates for psychotherapy. They are generally pessimistic about being helped and, as a result, might be argumentative with professional staff (confirm/disconfirm this by checking the TRT/Negative Treatment Indicators content scale). They need to be assured that they have been well understood and that their symptoms will not be ignored. Often framing interventions with biomedical terminology may make interventions more a acceptable (e.g., biofeedback procedures might be described as "neurological retraining").

MMPI-A Considerations The preceding descriptors are relevant for adolescent profiles. However, they also suggest school-related difficulties. Girls are likely to experience family problems (marital disagreements, financial concerns) and eating disorders. However, elevations on this scale are relatively rare among adolescents.

Low Scores on Scale 1

Low scores on females (but not males) suggest an absence of physical complaints and health-related concerns. They might also be generally alert, capable, responsible, and conscientious perhaps even to the point of being moralistic.

Scale 2. Depression (*D*)

Scale 2 comprises 60 items on the MMPI and 57 items on the MMPI-2. The Harris-Lingoes recategorization of the scale content suggests that these items are organized around the areas of brooding, physical slowness, subjective feelings of depression, mental apathy, and physical malfunctioning. Thus, high scores may indicate difficulties in

one or more of these areas. Patients seeking impatient psychiatric treatment are most likely to have Scale 2 as the highest point on their profiles. As would be expected, elevations on 2 typically decrease after successful psychotherapy. The relative elevation on Scale 2 is the single best predictor of a person's level of satisfaction, sense of security, and degree of comfort. Persons who score high on 2 are usually described as self-critical, withdrawn, aloof, silent, and retiring.

Any interpretation of scores on 2 needs to take into account the person's age and the implications of possible elevations on other scales. Adolescents typically score 5 to 10 points lower than nonpatient adults. In contrast, elderly persons usually score 5 to 10 points higher. A frequent pattern often referred to as the "neurotic triad" occurs when 1, 2, and 3 are all elevated. This suggests that the person has a wide variety of complaints, including not only depression, but also somatic complaints, irritability, difficulties with interpersonal relationships, work-related problems, and general dissatisfaction (see code types 12/21, 13/31, 23/32). An accompanying elevation on Scale 7 suggests that the self-criticalness and intropunitiveness of the depression also includes tension and nervousness. In some ways, moderate elevations on Scales 2 and 7 are a favorable sign for psychotherapy. Such elevations indicate that persons are motivated to change due to the discomfort they experience and they are also likely to be introspective and self-aware. Scales 2 and 7 are often referred to as the "distress scales" since they provide an index of the degree of personal pain, anxiety, and discomfort the person is experiencing (see code types 27/72). If an elevation on Scale 2 is accompanied by an elevation on Scale 8, it suggests that the depression is characterized by unusual thoughts, disaffiliation, isolation, and a sense of alienation (see code type 28/82).

High Scores on Scale 2

Moderate elevations on 2 might suggest a reactive depression, particularly if 2 is the only high point. The person would likely be confronting his or her difficulties with a sense of pessimism, helplessness, and hopelessness. These may even be characteristic personality features that become exaggerated when the person is confronted by current problems. He or she may have a sense of inadequacy, poor morale, and difficulty concentrating, which may be severe enough to create difficulties in working effectively. The person might be described as retiring, shy, aloof, timid, inhibited but also irritable, high-strung, and impatient. Since such persons are highly sensitive to criticism, they might attempt to avoid confrontations at all costs. This may result in their avoiding interpersonal relationships in general. Others may perceive them as being conventional, cautious, passive, and unassertive. Increasingly higher scores on 2 indicate an exaggeration of these trends. They may worry excessively over even minor problems, and their ability to deal effectively with interpersonal problems could be impaired. Their sense of discouragement might result in psychomotor retardation, lethargy, and withdrawal. Often they have a variety of somatic complaints (check for corresponding elevations on 1, HEA, and Harris-Lingoes D3/Physical Malfunctioning). In particular this might include feeling sluggish, tense, and having a low level of energy. They are also likely to have a preoccupation with death and suicide. Decisions may need to be made to determine whether they would require inpatient or outpatient treatment. Such decisions center on determining whether or not these clients are a danger to themselves.

Treatment Implications An important consideration is whether external (reactive) or internal (endogenous) factors are responsible for the depression. In addition, clients should be further assessed to determine the relative contribution of cognitions, social support, and the prevalence of vegetative symptoms. Treatment should be planned taking these considerations into account. An elevation on 2 raises the possibility of suicide. This is particularly true if the elevations are high to extremely high and if there are corresponding elevations on 4, 7, 8, and/or 9. Even though these elevations might raise the possibility of suicide, no clear "suicidal profile" has been found to be an accurate predictor. Any suggestion of suicidal behavior on the profile should be investigated further through a careful assessment of additional relevant variables (demographics, presence, clarity, lethality of plan, etc.). More specific information might also be obtained by noting relevant critical items (see Appendix K on p. 698) and then discussing these with the client. The presence of a moderate level of depression can be a positive sign for psychotherapy since they are likely to be highly motivated (but check possible negative indicators with elevations on TRT/Negative Treatment Indicators, *L, K,* and 1). In contrast, an extremely high score suggests they may be too depressed to experience sufficient motivation to change.

MMPI-A Considerations The preceding MMPI-2 descriptors and use of the Harris-Lingoes scales are also relevant for adolescents, particularly for girls. In addition, high adolescent scores on 2 suggest school-related difficulties (check A-sch/School Problems content scale) and a worsening of arguments between their parents (check A-fam/Family Problems content scale). They are less likely to act out but more likely to report eating problems (especially females), somatic complaints, and low self-esteem. Interpersonally, they will be introverted with a few friends.

Low Scores on Scale 2

Low scores generally indicate not only an absence of depression, but that the person is likely to be cheerful, optimistic, alert, active, and spontaneous. They report few difficulties trying to sleep, and do not appear to worry over their health. They may also be undercontrolled, self-seeking, and even prone to self-display. Males are likely to be confident, curious, able to easily make decisions, and relatively unconcerned what others think of them. Females are described as being cheerful, report few physical ailments, and unlikely to worry or become nervous. Sometimes a low 2 score might indicate a person who is denying significant levels of underlying depression.

Scale 3. Hysteria (*Hy*)

Scale 3 was originally designed to identify patients who had developed a psychogenically based sensory or motor disorder. The 60 items primarily involve specific physical complaints and a defensive denial of emotional or interpersonal difficulties. The types of physical complaints are generally quite specific and include areas such as fitful sleep, nausea, vomiting, headaches, and heart or chest pains (check HEA/Health Concerns scale). The important feature of persons who score high on this scale is that they simultaneously report specific physical complaints but also use a style of denial in which they may even express an exaggerated degree of optimism. One of the important

and primary ways in which they deal with anxiety and conflict is to channel or convert these difficulties onto the body. Thus, their physical complaints serve as an indirect expression of these conflicts. Their traits might be consistent with a histrionic personality in that they will demand affection and social support but do so in an indirect and manipulative manner. They are also likely to be socially uninhibited and highly visible. They can easily initiate relationships, yet their relationships are likely to be superficial. They will approach others in a self-centered and naive manner. They might act out sexually or aggressively, but have a convenient lack of insight into either their underlying motives or their impact on others. However, Scale 3 is quite heterogeneous in its item composition. The Harris-Lingoes item analysis has divided these into denial of social anxiety, need for affection, lassitude-malaise, somatic complaints, and inhibition of aggression. If Scale 3 is clearly elevated and a clinician is unclear regarding the meaning of the elevation, it can often be useful to formally score the Harris-Lingoes subscales (see section on Harris-Lingoes subscales).

An elevated Scale 3 is frequently found, with corresponding elevations on Scales 1 and 2 (see code types 12/21, 13/31, and 23/32). If *K* is also elevated, the person is likely to be inhibited, affiliative, overconventional, and to have an exaggerated need to be liked and approved of by others. This is particularly true if scales *F* and 8 are also low. A high score on 3 reduces the likelihood the person will be psychotic, even though Scales 6 and 8 might be relatively high.

High Scores on Scale 3

High scorers are likely to have specific functionally related somatic complaints. They will use a combination of denial and dissociation. They usually experience low levels of anxiety, tension, and depression and rarely report serious psychopathology such as hallucinations, delusions, and suspiciousness. Their insight about their behavior will be low as they both deny difficulties and have a strong need to see themselves in a favorable light. Persons with moderate scores, especially if educated and from higher socioeconomic groups, may have good levels of adjustment. Scale 3 might also be somewhat elevated in persons wishing to present a favorable impression for employment, thereby reflecting the endorsement of specific items denying any abnormality. With increasing scores, however, there is an exaggeration of denial, somatization, dissociation, immaturity, and low levels of insight. In particular, they may be perceived as highly conforming, immature, naive, childishly self-centered, and impulsive. They will have strong needs for approval, support, and affection but will attempt to obtain these through indirect and manipulative means. Thus, they are interpersonally indirect, with difficulty expressing hostility and resentment. Often, they will communicate with others to create an impact rather than to convey specific information. They will perceive events globally rather than attend to specific and often relevant details of a situation. Their physical difficulties typically worsen in response to increases in stress levels. The complaints can be either quite vague or quite specific and are of unknown origins. When their level of stress decreases, their physical difficulties will be likely to disappear very quickly. This is particularly true for persons with *T* scores over 80.

Treatment Implications Initial response to therapy is likely to be enthusiastic and optimistic, at least in part because the clients have strong needs to be liked. However,

they will be slow to gain insight into the underlying motives for their behavior because they use extensive denial and repression, typically denying the presence of any psychological problems. Often they will look for simplistic, concrete, and naive solutions to their problems. They would be likely to attempt to manipulate the therapist into a supportive and nonconfrontive role. As their defenses become challenged, they might become more manipulative, perhaps resorting to complaints of mistreatment and not being understood. At times, they may even become verbally aggressive. Their core conflicts are usually centered on issues of dependence versus independence. Often direct suggestion focusing on short-term goals is effective in creating change.

MMPI-A Considerations While the interpretations for adults with elevated Scale 3 can also be made with adolescents, they should be done with caution because of questions related to questionable validity with this population. In particular, the Harris-Lingoes subscales can help to clarify the meanings of scale elevations. Females (but not males) are still likely to have somatic complaints in response to stress. Males were more likely to have both school problems (check A-sch/School Problems content scale) and a history of suicidal ideation and gestures. However, Scale 3 is rarely a high point among adolescent males.

Low Scores on Scale 3

Low scores might be consistent with persons who are narrow-minded, cynical, socially isolated, conventional, constricted, and controlled. They might also have a difficult time trusting others and be difficult to get to know. Males are likely to be shy and more likely to report being worn out.

Scale 4. Psychopathic Deviate (*Pd*)

The purpose of Scale 4 is to assess the person's general level of social adjustment. The questions deal with areas such as degree of alienation from family, social imperviousness, difficulties with school and authority figures, and alienation from self and society (see ANG/Anger and FAM/Family Problems content scales). The original purpose of the scale was to distinguish those persons who had continuing legal difficulties, yet were of normal intelligence and did not report having experienced cultural deprivation. They were people who seemed unconcerned about the social consequences of their behavior and yet did not appear to suffer from neurotic or psychotic difficulties. An important rationale for developing the scale is that high scorers might not be engaged in acting out at the time of testing. In fact, they may often make an initial good impression, which could sometimes be described as "charming." Recent friends and acquaintances may not believe that they could even be capable of antisocial behavior. However, under stress or when confronted with a situation that demands consistent, responsible behavior, they would be expected to act out in antisocial ways. Even though they might get caught, these persons would still have a difficult time learning from their mistakes.

Different relatively normal groups will often have somewhat elevated Scale 4 profiles. This might include counterculture groups, which reflect their relative disregard for the values and beliefs of mainstream culture. Similarly, African Americans often score

higher, which might reflect their feelings that many of the rules and laws of the dominant culture are unfair and serve to disadvantage them. Normal persons who are graduate students in the humanities and social sciences often have somewhat elevated scores. More positive characteristics to be found with moderate elevations include frankness, deliberateness, assertion, sociability, and individualism. In addition, normal persons who are extraverted, risk takers, and have unconventional lifestyles (skydivers, police officers, actors) are also likely to have somewhat elevated Scale 4 profiles.

Relating Scale 4 with other corresponding scales can help to make more precise and accurate interpretations. If Scales 4 and 9 are elevated, it indicates that the persons not only have an underlying sense of anger and impulsiveness, but also have the energy to act on these feelings (see 49/94 code type and ASP/Antisocial Practices and ANG/Anger content scales). Their histories will almost always reveal extensive impulsive behavior. This acting out has frequently been done in such a way as to damage their family's reputation. They may also have been involved in criminal activity. However, moderate elevations on Scales 4 and 9 might suggest that the behaviors were less extreme and the person may have even been able to develop a good level of adjustment (see 49/94 code type). A psychotic expression of antisocial behavior might be consistent with elevations on both 4 and 8 (see 48/84 code type). A high 4 accompanied by a high 3 suggests that antisocial behavior might be expressed in covert or disguised methods or that the person might even manipulate another person into acting out for him or her (see 34/43 code type). Elevations on Scales 4 and 2 suggest that the person has been caught performing antisocial behavior and is feeling temporary guilt and remorse for his or her behaviors (see 24/42 code type).

High Scores on Scale 4

High scorers typically have problems with persons in authority, frequent marital and work difficulties, and poor tolerance for boredom. They can be described as having an angry disidentification with their family, society, or both. Although they might have been frequently caught for episodes of acting out, they are slow to learn from the consequences of their behavior. When confronted with the consequences of their actions, these individuals may feel genuine remorse, but this is usually short-lived. Their difficulty in profiting from experience also extends to difficulties in benefiting from psychotherapy. Usually, their relationships are shallow and characterized by recurrent turmoil, and they have difficulty forming any long-term loyalties. They are likely to blame others, particularly their families, when things go wrong. Others often perceive these individuals as angry, alienated, impulsive, and rebellious (see ASP/Antisocial Practices content scale) but also outgoing, extraverted, talkative, active, and self-centered. They usually have a history of involvement with the legal system as well as extensive alcohol or drug abuse. Because they resent rules and regulations, they will also have a history of work-related difficulties. Although they may often make an initial good impression, eventually they will have an outbreak of irresponsible, untrustworthy, and antisocial behavior.

Extremely high scorers might be aggressive or even assaultive. In addition, they will be unstable, irresponsible, self-centered, and most will have encountered legal difficulties because of their antisocial behaviors. In contrast, persons scoring in the moderate

ranges might be described as adventurous, pleasure-seeking, sociable, self-confident, assertive, unreliable, resentful, and imaginative.

Treatment Implications Because persons scoring high on 4 are usually verbally fluent, energetic, and intelligent, they might initially be perceived as good candidates for psychotherapy. However, their underlying hostility, impulsiveness, and feelings of alienation eventually surface. They are also likely to blame others for the problems they have encountered. As a result, they will eventually resist therapy and terminate as soon as possible (see TRT/Negative Treatment Indicators content scale). Part of their resistance is that they have difficulty committing themselves to any, including the therapeutic, relationship. If they are not feeling any subjective distress (low Scales 2 and 7), their motivation for change is likely to be particularly low. In addition, their original motivation for therapy may not have been to actually change, but rather to avoid some form of punishment, such as jail. Thus, their long-term prognosis in therapy is poor. Short-term goals that focus on documenting clear behavior change (rather than merely verbalizing it) would be indicated. Some sort of external motivation for therapy (e.g., condition of parole or continued employment) might also increase the likelihood that they will follow through on treatment.

MMPI-A Considerations Adolescents frequently have elevations on Scale 4, and it will be their highest overall scale. A full one third of the clinical sample used in the development of the MMPI-A had elevations of 65 or more. These generally high scores most likely reflect their often turbulent attempts to form a sense of identity and achieve independence from their parents. Thus, the elevation might be part of a temporary phase of development rather than a permanent enduring trait. However, high or extremely high scores will still reflect significant levels of pathology. Such scores are associated with delinquents who commit antisocial acts (see A-ang/Anger and A-con/Conduct Problems scales), are in conflict with their families (see A-fam/Family Problems), school-related difficulties (see A-Sch/School Problems), and are involved with drugs and/or alcohol (see MAC-R, ACK/Alcohol Drug Acknowledgment, and PRO/Alcohol Proneness supplementary scales). Often they report little guilt for this acting out and appear impervious to punishment. Additional difficulties might include externalizing behavior problems (lying, cheating, stealing, temper outbursts, jealousy) and school dropout. Boys frequently report physical abuse and having run away and girls similarly report physical abuse but also having been sexually abused. They are also likely to be sexually active. Often they are not particularly motivated to become involved in therapy. Because Scale 4 is quite heterogeneous with a correspondingly high number of descriptors, a formal scoring and inspection of the Harris-Lingoes scales can often be extremely useful in determining which of the scale descriptors is most appropriate.

Low Scores on Scale 4

Scores below 45 reflect persons who are overcontrolled, self-critical, rigid, conventional, and overidentified with social status. They might also be balanced, cheerful, persistent, and modest, but are somewhat passive and have difficulties asserting themselves. Males are less likely to become annoyed, less likely to resent being told what to

do, less likely to be moody, and rarely report having been in trouble with the law. Females are more likely to be cooperative, pleasant, and relaxed but less likely to have the following characteristics: temper tantrums, resent being told what to do, act stubbornly, feel that others don't care about them, argue about minor things, be envious, talk back to others, nag, be overly sensitive, be irritable, and tell people off for having made mistakes.

Scale 5. Masculinity-Femininity (*Mf*)

This scale was originally designed to identify males who were having difficulty with homosexual feelings and gender-identity confusion. However, it has been largely unsuccessful because a high score does not seem to clearly and necessarily relate to a person's sexual preference. Instead, it relates to the degree to which a person endorses items related to traditional masculine or feminine roles or interests. A high score for males is also positively correlated with intelligence and education. Thus, males who have completed university degrees will usually score an average of 5 *T* scores above ($T = 60–65$) the standardization sample; and those with less than a high school education will score, on average, 5 *T* scores lower. Interpretations, therefore, should consider the influence of education. The item content seems to be organized around the following five dimensions: personal and emotional stability, sexual identification, altruism, feminine occupational identification, and denial of masculine occupations. The items are scored in the opposite direction for females. Thus, high scores for males have traditionally been used to suggest a nonidentification with stereotyped masculine roles, whereas a high score for females has traditionally been used to suggest an identification with these masculine roles.

Since the original development of Scale 5, considerable change has occurred in society regarding the roles and behaviors of males and females. This, as well as other factors in scale construction, has caused some challenges to the current validity of Scale 5. Despite these challenges, the most recent study on college students indicated that the behavioral correlates now are quite similar to what they were a generation ago (A. Todd & Gynther, 1988). However, there is insufficient research to fully understand the behavioral correlates of Scale 5, and there are still unanswered questions about its original development. These concerns were sufficient to cause the restandardization committee to consider deleting it from the MMPI-2 and MMPI-A (Butcher & Williams, 1992; C. Williams & Butcher, 1989b). The practical implication is that clinicians must make Scale 5 interpretations quite cautiously, particularly for females.

An important consideration regarding Scale 5 is that it is not an actual clinical scale in the same sense as most of the other scales. It does not actually assess any pathological syndromes and thus does not provide clinical information. However, it can be useful in providing color or tone to the other scales. This means interpretations should first be made of the other scales and then the meaning of the relative score on Scale 5 should be taken into consideration. For example, an elevation on Scale 4 (*Pd*) would indicate that the person is impulsive, might act out under stress, and feels alienated from his or her self or society. If the person scoring a high 5 is a male and also scores low on Scale 4, he would be likely to express his dissatisfaction through action, have low insight into his behavior, and place emphasis on physical strength. In contrast, a high scale on 4

accompanied by an elevated 5 suggests that the person will be more introspective, sensitive, articulate, and may channel his or her antisocial feelings toward creating social change. However, in deciding what should be considered to be a high or low 5, the person's level of education and socioeconomic status should always be considered.

A high score on 5 for males should never be used to diagnose homosexuality. Instead, high-scoring males are more likely to suggest that the person has aesthetic interests, sensitivity to others, a wide range of interests, tolerance, passivity, and is capable of expressing warmth. Males with moderate scores would be inner directed, curious, clever, and have good judgment and common sense. In some cases, extremely high scores might suggest homosexuality, but this is raised only as a possibility: The scores themselves are not diagnostic. The scale is also quite susceptible to faking because the meanings of the items are fairly transparent. Thus, a person wishing to conceal his or her gender-identity confusion could easily alter the items. In contrast, males with low scores could be expected to endorse traditional masculine interests and be described as easygoing, adventurous, but also sometimes coarse.

Females who lack much education and score low on Scale 5 are usually described as fulfilling traditionally feminine roles and occupations. They will be passive, submissive, modest, sensitive, and yielding. In contrast, highly educated females who score in the same ranges are likely to be intelligent, forceful, considerate, insightful, conscientious, and capable. Females who score high are more likely to be involved in more traditionally male roles and occupational areas such as mechanics and science. They are frequently described as aggressive, competitive, confident, adventurous, and dominating.

Sometimes, males have elevations on both 4 and 5. Such a profile is likely to reflect a person who is not only unconventional, but will also be flamboyant about expressing this unconventionality. Thus, he may enjoy openly defying and challenging conventional modes of appearance and behavior (see 45/54 code type). In contrast, males who score low on 5 and high on 4 will be likely to make an obvious, perhaps even compulsive, display of their masculinity. Females with a high 5 and 4 will rebel against traditional expressions of femininity. As the elevation on 4 becomes progressively higher, this rebellion is likely to become correspondingly more deviant. Females having a profile with a high 4 and low 5 will similarly feel angry and hostile but will find it very difficult to express these feelings. This is likely to produce a considerable inner tension and turmoil. Sometimes, males have high scores on 5 with corresponding high scores on 2 and 7, and occasionally on 4. These males will present themselves as self-effacing, weak, submissive, inferior, and guilty. In high school or college, they may have taken on either the role of the self-critical school clown or a withdrawn "egghead." Females expressing these qualities would be likely to have the same pattern of scores, except they would have a valley rather than an elevation on 5.

High Scores on Scale 5 (Males)

High scores for males have traditionally been interpreted as suggesting that they are likely to be undemanding, shy, emotional, curious, and creative, with a wide range of intellectual interests. Extremely high scores might suggest males who are not only involved in traditional feminine interests but also have gender-identity confusion. They might be effeminate, passive, experience homoerotic feelings, and have difficulty asserting themselves. Because of their passivity, they may experience marital problems

as they have difficulty assertively fulfilling their partners' needs. In rare cases, some high-scoring males might develop a reaction formation against their passivity and display an exaggerated expression of masculinity. Thus, they would be outwardly similar to low-scoring males on Scale 5 although their inner experience of these behaviors would be quite different. In contrast, moderate scorers will be expressive, demonstrative, empathic, individualistic, express interpersonal sensitivity, have aesthetic interests, be self-controlled, nurturant, and exercise good common sense. They should be able to communicate their ideas effectively and are psychologically sophisticated, idealistic, and inner directed. College and seminary students usually score in the moderate ranges.

Treatment Implications Moderate scores suggest that the individual is insightful, sensitive, and introspective, all of which are qualities conducive to psychotherapy. A high 5 reduces the likelihood that any existing pathology will be acted out. However, with increasing elevations, there are likely to be important issues related to passivity, dependency, impracticality, problems dealing with anger, and difficulties with heterosexual adjustment. In contrast, low-scoring men may have difficulties with psychotherapy because they have low verbal skills, are not psychologically minded, have a narrow range of interests, and tend to be action oriented. The most successful approaches for working with low-scoring males involve clear behavioral descriptions with concrete, short-term goals.

MMPI-A Considerations Scale 5 elevations for males were rare on both the MMPI-A clinical and normative samples. While many of the descriptions for high-scoring adult males should be considered for adolescents, any interpretations should be made with caution. However, they will seem interested in stereotypically feminine interests, deny stereotypically masculine interests, and are less likely to act out. If there are correspondingly high elevations on other scales suggesting acting out (Scales 4, 9, *F*), these should be given more consideration than the suppression value of an elevated Scale 5.

Low Scores on Scale 5 (Males)

Low-scoring males will be domineering and impersonal. Their interests might be somewhat narrow, and they will lack originality. They will show little interest in understanding the underlying motivation behind their own behavior or the behavior of others. In contrast, they will prefer action over contemplation and will be generally uncomfortable with their feelings. They will place considerable emphasis on athletic prowess and physical strength. Thus, they will act in a traditionally masculine manner. They might also be described as self-indulgent, independent, and narcissistic. Sometimes their masculine strivings might even be overdone and expressed in a rigid, almost compulsive manner. Extremely low scorers might even be expressing exaggerated masculine behavior to conceal serious doubts regarding their own masculinity.

High Scores on Scale 5 (Females)

Because the scoring for females is reversed, a high score will mean the opposite for females as it would for males. Thus, high-scoring females would be endorsing traditionally masculine interests and activities, and may be involved in occupations that are

more frequently occupied by males. They will often be described as confident, spontaneous, bold, unsympathetic, competitive, logical, and unemotional. When compared with low-scoring females, their physical health is likely to be better; and they will more frequently be involved in active hobbies and interests. As the scale elevation increases, they correspondingly might be more aggressive, tough-minded, and domineering. They may often be rebelling against the traditional female role and feel uncomfortable in heterosexual relations.

Treatment Implications High-scoring females might be difficult to engage in traditional psychotherapy because they usually do not value introspection and insight and have difficulty articulating their problems and expressing emotions. In contrast, low-scoring females who are well educated will be articulate, sensitive, have a wide range of interests, and can clearly express their emotions, thus making them good candidates for psychotherapy. However, low-scoring females with low levels of education might be difficult to work with because they are likely to be extremely passive, superficially compliant, but have difficulty actually implementing change.

MMPI-A Considerations Further research needs to be conducted on the behavioral correlates of both high- and low-scoring adolescent females. However, tentative interpretations would indicate that high-scoring females have stereotypically masculine interests.

Low Scores on Scale 5 (Females)

College-educated females with low scores on 5 will be tender; emotional; have a balanced view of gender-role behavior; express aesthetic interests; and be capable, competent, and conscientious. They may still endorse many traditionally feminine roles and behaviors. They are more likely to have a greater number of health-related complaints than high-scoring females, and their hobbies and interests will be more passive. In contrast to low-scoring educated females, low-scoring females with limited education are typically described quite differently. They may be caricatures of traditionally feminine behavior. They are likely to be modest, passive, constricted, and yielding. They may attempt to make others feel guilty by taking on an excessive number of burdens. As a result, they might be complaining and self-pitying, and might spend time finding faults in others. It is not unusual to have a low 5 accompanied by elevations on the "neurotic triad" (Scales 1, 2, and 3). A low 5 accompanied by elevations on Scales 4 and 6 has been referred to as the "Scarlett O'Hara" profile because the person is likely to express an exaggerated degree of femininity combined with covert manipulations, underlying antisocial feelings, and hypersensitivity.

Scale 6. Paranoia (*Pa*)

Scale 6 was designed to identify persons with paranoid conditions or paranoid states. It measures a person's degree of interpersonal sensitivity, self-righteousness, and suspiciousness. Many of the 40 items center on areas such as ideas of reference, delusional beliefs, pervasive suspiciousness, feelings of persecution, grandiose self-beliefs, and interpersonal rigidity. Whereas some of the items deal with overt psychotic content,

other less extreme questions ask information related to perceived ulterior motives. The Harris-Lingoes subscales divide the items in Scale 6 into ideas of external influence, poignancy (feelings of being high-strung, sensitive, having stronger feelings than others, and a sense of interpersonal distance), and naiveté (overly optimistic, high morality, denial of hostility, overly trusting, and vulnerability to being hurt).

Mild elevations on Scale 6 suggest that the person is emotional, soft-hearted, and experiences interpersonal sensitivity. As the elevation increases, a person's suspicion and sensitivity become progressively more extreme and consistent with psychotic processes. He or she may have delusions, ideas of self-reference, a grandiose self-concept, and disordered thought processes. In contrast, low-scoring persons are seen as being quite balanced. However, there are some differences between the descriptions given for low-scoring males as opposed to low-scoring females. Low-scoring males are described as cheerful, decisive, self-centered, lacking in a strong sense of conscience, and having a narrow range of interests. Females are somewhat differently described as mature and reasonable. Persons scoring extremely low might actually be paranoid but are attempting to hide their paranoid processes. Thus, they would actually have characteristics quite similar to high-scoring persons.

In some ways, Scale 6 is quite accurate in that high-scoring persons will usually have significant levels of paranoia. However, the contents of most of the 40 items are fairly obvious. Thus, a person wanting to conceal his or her paranoia, because of fear over the imagined consequences of detection, could do so quite easily. This means it might be possible for low or moderate scores to still be consistent with paranoia. This is especially true for bright and psychologically sophisticated persons. They might mask their paranoia not only on the test, but also in real life. Thus, they might be a member of some extreme political group or religious cult that provides some degree of social support for their underlying paranoid processes. However, if the scale is clearly elevated, it is an excellent indication of paranoia.

A pronounced elevation on Scales 6 and 8 is highly suggestive of paranoid schizophrenia, regardless of the elevations on the other scales (see 68/86 code type).

Another frequent combination is a corresponding elevation on Scale 3. Such persons would then repress their hostile and aggressive feelings and appear naive, positive, and accepting. They might easily enter into superficial relationships, but after these relationships deepen, their underlying suspiciousness, hostility, ruthlessness, and egocentricity would become more openly expressed (see 36/63 code type).

High Scores on Scale 6

Extremely high scores on Scale 6 indicate persons who are highly suspicious, vengeful, brooding, resentful, and angry. They will feel mistreated and typically misinterpret the motives of others, feeling that they have not received a fair deal in life. They are likely to have a thought disorder with accompanying ideas of reference, delusional thinking, fixed obsessions, compulsions, and phobias. Their thinking will be extremely rigid, and they will be quite argumentative.

Moderate elevations are much less likely to reflect overtly psychotic trends. However, these persons are still likely to be suspicious, argumentative, potentially hostile, and quite sensitive in interpersonal relationships. They might easily misinterpret the benign statements of others as personal criticisms. They would then enlarge on and

brood over these partially or wholly invented criticisms. It would be difficult to discuss emotional problems with them. Although they might have underlying feelings of anger, they are likely to deny these in a rigidly moralistic and intellectual manner. They would also be likely to defend themselves from anxiety through intellectualization and would use projection to deny underlying feelings of hostility. They might then express their own hostility through indirect means and yet appear outwardly self-punishing. They will feel as if they have gotten an unfair deal from life and feel particular resentment toward family members.

Persons with mild elevations on 6 are usually described in relatively favorable terms. These include hardworking, industrious, moralistic, sentimental, softhearted, peaceable, generous, and trusting unless betrayed. They are also likely to be intelligent, poised, rational, fair-minded, and have a broad range of interests. However, they might also tend to be submissive, prone to worry, high strung, dependent, and lacking in self-confidence. The preceding descriptions are particularly likely among nonpatient groups. Psychiatric patients with the same elevations are described somewhat differently as being oversensitive, slightly paranoid, suspicious, and feeling as if their environment is not sufficiently supportive.

Treatment Implications Considering the relative elevation of this scale can be extremely important because it provides an index of the degree to which clients can develop a trusting relationship, their attitudes toward authority figures, and their degree of flexibility. Very high scores suggest psychotic processes possibly requiring medication (check BIZ/Bizarre Mentation and critical item clusters related to mental confusion and persecutory ideas subscales). Psychotherapy with individuals scoring high on Scale 6 would be extremely difficult because of their rigidity, poor level of insight, and suspiciousness (check TRT/Negative Treatment Indicators Content scale). In addition, they do not like to discuss emotional issues, overvalue rationality, and are likely to blame others for their difficulties. They frequently will not return following the initial session and leave feeling that they have not been understood. Thus, a major challenge with an intake is to make sure that they feel understood. During subsequent sessions, they might attempt to manipulate the therapist by implicitly suggesting they will terminate. They are often argumentative, cynical, and resentful, thereby making it difficult to establish a relationship of mutual trust, empathy, and respect. If brooding and resentment are particularly pronounced, it might be important to assess for potential dangerousness toward others.

MMPI-A Considerations Elevations are consistent with academic problems including poor grades and suspension (check A-sch/School Problems content scale). Clinical girls report significant disagreements with their parents (check A-fam/Family Problems content scale). Clinical boys are described as hostile, withdrawn, immature, and argumentative; they feel persecuted and are not well liked by their peers. In addition, they are perceived as being overly dependent on adults, attention-seeking, resentful, anxious, and obsessed; they feel as if they are bad and deserving of punishment. Because the items between the MMPI, MMPI-2, and MMPI-A are the same, the Harris-Lingoes scales can be used to understand possible patterns of item endorsement.

Low Scores on Scale 6

Most persons with low scores on 6 are described as being quite balanced. Males tend to be cheerful, decisive, lacking in a sense of conscience, self-centered, effectively in control of their emotions, and have a narrow range of interests. Females are somewhat more favorably described as being not only balanced, but also mature and reasonable. Both males and females are likely to be able to accept the challenges of life and are trusting, loyal, decisive, and self-controlled. Whereas the preceding descriptions tend to be true for nonpatient groups, persons having the same scores among patient groups are described as oversensitive, uninsightful, introverted, undependable, touchy, and rough, with poorly developed consciences and a narrow range of interests. Persons with extremely low scores might be paranoids who are attempting to hide their thought processes. Thus, they would be similar to high-scoring persons and could be described as guarded, defensive, shy, secretive, evasive, and withdrawn. This would be particularly likely if the following conditions are present: The validity scales indicate a defensive pattern of underreporting psychopathology (*L* and *K* are above 60 and above *F*), at least one of the *T* scores on one of the other clinical scales is above 65, and Scale 1 is less than 35 and is the lowest on the profile.

Scale 7. Psychasthenia (*Pt*)

The 48 items on Scale 7 were originally designed to measure the syndrome of psychasthenia. Although psychasthenia is no longer used as a diagnosis, it was current when the MMPI was originally developed. It consisted of compulsions, obsessions, unreasonable fears, and excessive doubts. Thus, it is quite similar to what today would be an anxiety disorder with obsessive-compulsive features. However, there are important differences. Scale 7 measures more overt fears and anxieties that the person might be experiencing (check also ANX/Anxiety). In contrast, persons having an obsessive-compulsive disorder could potentially score quite low on 7 because their behaviors and obsessions are effective in reducing their levels of anxiety (check the OBS/Obsessiveness content scale). Although an elevation on Scale 7 may suggest the possibility of an obsessive-compulsive disorder, other anxiety-related disorders or situational states could also produce an elevation.

Scale 7 is the clinical scale that most clearly measures anxiety and ruminative self-doubt. Thus, along with elevations on Scale 2, it is a good general indicator of the degree of distress the person is undergoing. High scorers are likely to be tense, indecisive, obsessionally worried, and have difficulty concentrating. In a medical context, they are prone to overreact to even minor medical complaints. They are usually rigid, agitated, fearful, and anxious. The most frequent complaints will be related to cardiac problems as well as difficulties related to their gastrointestinal or genitourinary systems. In nonmedical and more normal populations, high scorers are likely to be high-strung, articulate, individualistic, and perfectionistic, with extremely high standards of morality.

If both Scales 7 and 2 are moderately elevated, it suggests a good prognosis for therapy because these individuals are sufficiently uncomfortable to be motivated to change. They are likely to stay in treatment longer than most other groups. Although

their progress will be slow, there is likely to be progressive improvement (see 27/72 code type). If their scores are extremely high, they might require antianxiety medication to enable them to relax sufficiently to be able to coherently discuss their difficulties. It might also be important to note the elevation of 7 in relationship to 8. If 7 is significantly higher than 8, it indicates the person is still anxious about and struggling with an underlying psychotic process. However, if 7 is quite low (10 *T* score points or more) in comparison to 8, the person is likely to have given up attempting to fight the disorder, and his or her psychotic processes are either of a chronic nature or likely to become more chronic (see 78/87 code type).

High Scores on Scale 7

Elevations on Scale 7 suggest persons who are apprehensive, worrying, perfectionistic, and tense, and who may have a wide variety of superstitious fears. Mild elevations suggest that, in addition to a certain level of anxiety, these persons will be orderly, conscientious, reliable, persistent, and organized, although they will also lack originality. Even minor problems might become a source of considerable concern. They will overreact and exaggerate the importance of events. Although they might attempt to use rationalization and intellectualization to reduce their anxiety, these defenses are rarely successful. With increasing elevations, they are likely to experience greater levels of self-doubt and be progressively more rigid, meticulous, apprehensive, uncertain, and indecisive. They are shy and experience social difficulties, frequently worrying about their degree of acceptance and popularity. They may be rigid and moralistic with high standards for themselves and others. Defenses against their anxiety could be expressed in a variety of rituals and difficulty in concentrating. They will be highly introspective, self-critical, self-conscious, and feel a generalized sense of guilt. Extremely high scores might indicate a disruption in a person's ability to perform daily activities.

Treatment Implications Because persons scoring high on Scale 7 experience clear, overt levels of discomfort, tension, and cognitive inefficiency, they are highly motivated to change. Thus, they will usually stay in therapy. Progress tends to be slow although steady. The immediate task is to work directly with their anxiety, which might be accomplished through procedures such as cognitive restructuring, hypnosis, relaxation, or systematic desensitization. For some clients, their anxiety may be sufficiently high that a short-term regimen of antianxiety medication might be considered to help them work more constructively in a therapeutic context and function in their daily activities. Insight-oriented therapy should be used with caution, as they will have a tendency to intellectualize and ruminate indefinitely without making any concrete changes. A further potential difficulty is that they may be overly perfectionistic and rigid, thereby making it difficult to either accept insights or to think in a flexible, problem-solving manner.

MMPI-A Considerations Few descriptors have been found for adolescents with high scores on *Pt* in part because, it is speculated, an early (adolescent) rigid personality style may not become problematic until later in adult life. Girls from clinical populations are likely to be depressed, may make suicidal threats, are more likely to steal, and report significant disagreements with their parents. Boys from clinical populations

are likely to have low self-confidence and may have been sexually abused. However, more research needs to be performed to more clearly understand the behavioral correlates of high Scale 7-scoring adolescents.

Low Scores on Scale 7

Low scorers are likely to be relaxed, warm, cheerful, friendly, alert, and self-confident. They will approach their world in a balanced manner and are often described as efficient, independent, placid, and secure. Both males and females are less likely to be critical of others and typically have few fears. Males are also less likely to worry about health concerns. Females are unlikely to be nervous, critical of themselves or others, have bad dreams, and worry over minor issues or events. It is rare for persons referred for treatment to have low scores on this scale.

Scale 8. Schizophrenia (*Sc*)

Scale 8 was originally designed to identify persons who were experiencing schizophrenic or schizophrenic-like conditions. This goal has been partially successful in that a diagnosis of schizophrenia is raised as a possibility in the case of persons who score extremely high. However, even persons scoring quite high would not necessarily fulfill the criteria for schizophrenia, in part because the items in the scale cover a highly diverse number of areas. Thus, elevations can occur for a variety of reasons, which means that the descriptions of high scorers would also be quite varied. The items assess areas such as social alienation, apathy, poor family relations, unusual thought processes, and peculiarities in perception. Other questions are intended to measure reduced efficiency, difficulties in concentration, general fears and worries, inability to cope, and difficulties with impulse control. Because of the many scale items, heterogeneity of their content, and the resulting numerous potential descriptors for individuals scoring high on Scale 8, it would be advisable to consult the Harris-Lingoes subscales to more fully understand the meanings of elevations. The following six different content areas have been described by Harris and Lingoes (1968):

1. Social alienation.
2. Emotional alienation.
3. Lack of ego mastery-cognitive (strange thought processes, fear of losing his or her mind, difficulty concentrating, feelings of unreality).
4. Lack of ego mastery-cognative (difficulty coping with everyday life, low interest in life, hopelessness, depression).
5. Lack of ego mastery-defective inhibition (impulsive, hyperactive, sense of being out of control, impulsive, laughing or crying spells).
6. Bizarre sensory experiences.

In general, an elevated score on Scale 8 suggests the person feels alienated, distant from social situations, and misunderstood. He or she might have a highly varied fantasy life and, when under stress, will withdraw further into fantasy. Others will most likely perceive the person as eccentric, seclusive, secretive, and inaccessible. He or

she will often have a difficult time maintaining a clear and coherent line of thought. Communication skills will be poor; often, other people will feel they are missing some important component of what this individual is trying to say. The person will typically not make clear and direct statements and often will have difficulty focusing on one idea for very long.

Age and race are important when deciding what should be considered a high versus a low score on Scale 8. Many populations of African Americans score somewhat higher on 8, which might reflect their greater level of social alienation and estrangement. However, this may be more related to education and socioeconomic status than ethnicity (Greene, 1987, 1991, 2000). Adolescents also score higher on Scale 8, which might be consistent with their greater openness to unusual experiences, turmoil in establishing a solid sense of identity, and greater feelings of alienation. Some groups of relatively normal persons might have mild elevations on 8. These might include individuals developing sensory impairments, persons with organic brain disorders, or unconventional persons who identify with the counterculture. Persons who have had a variety of drug experiences may score somewhat higher on 8. This may reflect the direct effects of the drugs themselves rather than suggest greater levels of pathology.

Simultaneous elevations on 4 and 8 indicate persons who feel extremely distrustful and alienated from their world. They perceive their environment as dangerous and are likely to react to others in a hostile and aggressive fashion (see 48/84 code type). Another important but unusual profile is an elevation on 8 along with an elevation on 9. Such persons will be likely to constantly deflect the direction of conversation, frequently diverting it to unusual tangents. They are likely to have not only a distorted view of their world, but also the energy to act on these distorted perceptions (see 89/98 code type). Another important pattern is the prognostic significance associated with the relative height of 7 and 8 (see 78/87 code type) and the schizoid profile of elevated *F,* 2, 4, 8, and 0 (see Scale 2).

High Scores on Scale 8

A high score suggests persons who have unusual beliefs, are unconventional, and may experience difficulties concentrating and focusing their attention. In moderately elevated protocols, they might merely be aloof, different, and approach tasks from an innovative perspective. They may have philosophical, religious, or abstract interests, and care little about concrete matters. Others might describe them as shy, aloof, and reserved. Progressively higher scores would be likely to reflect individuals with greater difficulties in organizing and directing their thoughts. They might have aggressive, resentful, and/or hostile feelings yet cannot express these feelings. At their best, they might be peaceable, generous, sentimental, sharp-witted, interesting, creative, and imaginative. Very high elevations suggest persons with bizarre mentation, delusions, highly eccentric behaviors, poor contact with reality, and possibly hallucinations (see BIZ/Bizarre Mentation content scale). They will feel incompetent, inadequate, and be plagued by a wide variety of sexual preoccupations, self-doubts, and unusual religious beliefs. However, extremely high scores rarely occur, even among diagnosed schizophrenics. These extremes are likely to reflect unusual experiences reported by unusually anxious patients, adolescent adjustment reactions, prepsychotics, borderline personalities, or relatively well-adjusted persons who are malingering.

Treatment Implications Because high-scoring persons have difficulty trusting others and developing relationships, therapy might be difficult, especially during its initial stages. However, such individuals tend to stay in therapy longer than many other types of clients and may eventually develop a relatively close and trusting client/therapist relationship. Because of the often-chronic nature of their difficulties, their prognosis is frequently poor. If their thought processes are extremely disorganized, referral for medication might be indicated.

MMPI-A Considerations Both boys and girls report a higher rate of having several school-related problems, with boys frequently being suspended and girls being unlikely to report having had any significant achievements (check A-sch/School Problems content scale). In addition, the possibility of sexual abuse should be investigated. Girls are likely to report increased disagreements with their parents (check A-fam/Family Problems content scale) and, among clinical populations, may be aggressive, threaten suicide, act out, and have outbursts of temper. In contrast, clinical boys are described as having behaviors such as being guilt-prone, shy, withdrawn, fearful, and perfectionistic; showing low self-esteem; being "clingy"; and having somatic complaints (nausea, headaches, dizziness, stomach pains). Clinical boys with quite high elevations might also have psychotic features including delusions, hallucinations, ideas of reference, grandiose beliefs, or peculiar speech and mannerisms (check A-biz/Bizarre Mentation content scale).

Low Scores on Scale 8

Persons scoring low are likely to be cheerful, good-natured, friendly, trustful, and adaptable. However, they are also likely to be overly accepting of authority, restrained, submissive, unimaginative, and avoid deep and involved relationships with others.

Scale 9. Hypomania (*Ma*)

The 46 items on Scale 9 were originally developed to identify persons experiencing hypomanic symptoms. These symptoms might include cyclical periods of euphoria, increased irritability, and excessive unproductive activity that might be used as a distraction to stave off an impending depression. Thus, the items are centered on topics such as energy level, irritability, egotism, and expansiveness. The Harris-Lingoes subscales classify the content of the items under amorality, psychomotor acceleration, imperturbability, and ego inflation. However, hypomania occurs in cycles. Thus, persons in the acute phase were unable to be tested because of the seriousness of their condition. Further, some persons might score quite low on Scale 9, which might reflect the depressive side of their cycle. These low scorers, then, might still develop a hypomanic state and may have actually been hypomanic in the past.

The scale is effective not only in identifying persons with moderate manic conditions (extreme manic patients would be untestable) but also in identifying characteristics of nonpatient groups. A full 10% to 15% of normals have elevations on this scale, suggesting characteristics such as an unusually high drive level. Males with moderate to mild elevations and with no history of psychiatric disturbance might be described as warm, enthusiastic, outgoing, and uninhibited. They would most likely be able to expend a considerable amount of energy over a sustained period of time. They might also be easily

offended, hyperactive, tense, and prone to periods of worry, anxiety, and depression. Others might describe them as expressive, individualistic, generous, and affectionate. Nonpatient females are likely to be frank, courageous, talkative, enthusiastic, idealistic, and versatile. Their husbands are likely to describe them as making big plans, wearing strange or unusual clothes, stirring up excitement, becoming very excited for no reason, being risk takers, and telling people off. High-scoring males were described by their wives as demanding excessive attention, being bossy, talking back to others without thinking, whining, and taking nonprescription drugs.

Age and race are important when evaluating what should be considered a high or low score. Some studies have indicated that certain populations of African Americans score higher than European Americans. Also, younger populations (adolescents and college-age students) score somewhat higher than nonpatient adults. In contrast, elderly persons often score quite low on Scale 9.

Useful information can often be obtained by interpreting the significance of corresponding scores on 9, 2, 7, and *K*. Usually 9 and 2 are negatively correlated. Sometimes, however, they may both be elevated, reflecting an agitated state in which the person is attempting to defend or distract himself or herself from underlying hostile and aggressive impulses. Sometimes such persons are highly introspective and narcissistically self-absorbed. Scales 9 and 2 can also be elevated for certain types of organically impaired patients. Profiles in which 2 and 7 are low (suggesting a minimum of psychological distress), combined with an elevation on 9, might be consistent with males who have an almost compulsive need to seek power and place themselves in narcissistically competitive situations. If the profile is accompanied by an elevation on *K,* these males are likely to be managerial, autocratic, and power hungry, and expend a considerable degree of effort in organizing others. Their self-esteem would often be dependent on eliciting submission and weakness from others. What they usually receive from others is a grudging deference rather than admiration. Females having this profile are likely to be prone to exhibitionistic self-display and to be extremely concerned with their physical attractiveness.

High Scores on Scale 9

Extremely high scores are suggestive of a moderate manic episode. These individuals will be maladaptively hyperactive and poorly focused, and will have flighty ideas, an inflated sense of self-importance, and low impulse control. They might make an initial good impression because they are enthusiastic, friendly, and pleasant. However, they also tend to be deceptive, manipulative, and unreliable, ultimately causing interpersonal difficulties. While others might perceive them as creative, enterprising, and ingenious, a high Scale 9 appraisal of what he or she can actually accomplish is unrealistic. Thus, these clients have an unwarranted sense of optimism. They are likely to become irritable at relatively minor interruptions and delays. Although they expend a considerable amount of energy, their activity usually will be unproductive because it is unfocused. Others might also perceive them as restless and agitated. They will quickly develop relationships with others, but these relationships will be superficial.

Persons with more moderate elevations are often more able to focus and direct their energy in productive directions. Nonpatients will be direct, energetic, enthusiastic,

sociable, independent, optimistic, and have a wide range of interests. They might also be somewhat guileful, overactive, impulsive, persuasive, and prefer action to thought. They sometimes show mood difficulties and experience elation without cause. They are also described as self-centered and impulsive. Sometimes scores alone are not sufficient to distinguish a person who is energetic, optimistic, and focused from a person who is scattered, ineffective, and hyperactive. Useful information might be obtained by noting relevant critical items, interpreting the Harris-Lingoes subscales, or integrating relevant historical information.

Treatment Implications Because elevation on Scale 9 indicates distractibility and overactivity, these clients might be difficult to work with. They might resist focusing on problems by diverging onto irrelevant tangents and object to psychological interpretations of their behavior. The client might attempt to persuade the therapist into believing in his or her grandiose plans for change but these are seldom followed through. They also tend to use denial and avoid self-examination. Because they have a low frustration for tolerance and become easily irritated, therapy sessions might become dramatic. Frequently, they show disregard for prearranged appointment times and cancel because they are too busy. This generally resistive stance suggests they might optimally benefit from non- or self-directive interventions or paradoxical strategies. They might need to be formally evaluated for the possibility of a bipolar disorder with follow-up for appropriate medication. Further assessment should also be made regarding alcohol or drug abuse (check MAC-R, AAS/Addiction Acknowledgment scale, and APS/Addiction Potential scale). Clients with low scores on Scale 9 are likely to be apathetic, depressed, fatigued, inadequate, and pessimistic. As a result, they might have difficulty becoming motivated and require a concrete action program with a high degree of structure.

MMPI-A Considerations Moderate elevations suggest that the individual is enthusiastic, animated, and takes an interest in things. However, higher elevations suggest underachievement in school and problems at home (check A-sch/School Problems and A-fam/Family Problems content scales). Scale 9 elevations might also reflect irrational, manic behaviors and antisocial acts (check A-con/Conduct Problems content scale). Among boys, amphetamine use is relatively common. They are typically insensitive to criticism, do not like to reflect on their behavior, and are, therefore, unmotivated to become involved in therapy. They may also believe that they know more than authority figures, and feel that such persons punish people unjustly. They might be self-confident, oppositional, take advantage of others, and deny any social discomfort.

Low Scores on Scale 9

Persons scoring low on Scale 9 are likely to have low levels of energy and activity. They are often described as dependable, responsible, conventional, practical, and reliable, but may also lack self-confidence. For example, they would be unlikely to dress in strange or unusual clothes, curse, stir up excitement, or talk back to others. They might be seclusive, withdrawn, quiet, modest, overcontrolled, and humble. Low scores are more frequently found among the elderly than among younger populations. Extremely low scores suggest serious depression, even if Scale 2 is in normal limits.

Scale 0. Social Introversion (*Si*)

This scale was developed from the responses of college students on questions relating to an introversion-extraversion continuum. It was validated based on the degree to which the students participated in social activities. High scores suggest that the respondent is shy, has limited social skills, feels uncomfortable in social interactions, and withdraws from many interpersonal situations. In particular, these individuals may feel uncomfortable around members of the opposite sex. They would prefer to be alone or with a few close friends than with a large group. One cluster of items deals with self-depreciation and neurotic maladjustment, whereas the other group deals with the degree to which the person participates in interpersonal interactions. The different item contents have been further organized around the areas of shyness/self-consciousness, social avoidance, and the extent that a person feels alienated from self and others (Ben-Porath, Hostetler, Butcher, & Graham, 1989). These contents form subscales that can be used in conjunction with the Harris-Lingoes subscales to help determine the different variables related to why a person had an elevation on Scale 0 (see section on Harris-Lingoes and *Si* subscales).

Scale 0 is similar to 5 in that it is used to "color" or provide a different emphasis to the other clinical scales. Thus, interpretations should first be made without considering 5 and 0 and, later, the implications of these scales should be included. As a result, code types involving 0 have not been included in the section on two-point codes. Elevations on 0 help provide information on the other scales by indicating how comfortable persons are with interactions, their degree of overt involvement with others, the effectiveness of their social skills (check SOD/Social Discomfort content scale), and the likelihood that they will have a well-developed social support system. A low score on 0 will often reduce the degree of pathology that might otherwise be suggested by elevations on the other scales. A low 0 also suggests that, even if persons have a certain level of pathology, they are able to find socially acceptable outlets for these difficulties. In contrast, a high 0 suggests an exaggeration of difficulties indicated by the other scales. This is particularly true if 0, 2, and 8 are all elevated. This suggests that the person feels socially alienated, withdrawn, is self-critical, and has unusual thoughts. However, he or she is not likely to have an adequate social support group to help in overcoming these difficulties. Although an elevated 0 can suggest an increase in personal difficulties, it often reflects a decreased likelihood of acting out. This is further supported by corresponding elevations on 2 and 5 (for males or a lowering for females). As a result, 0, 2, and 5 are often referred to as *inhibitory scales.*

High Scores on Scale 0

Persons scoring high on Scale 0 will feel uncomfortable in group interactions and may have poorly developed social skills. They may be self-effacing, lacking in self-confidence, submissive, shy, and timid. Others might experience them as cold, distant, rigid, and difficult to get to know. Extremely high scorers are described as withdrawn, ruminative, indecisive, insecure, and retiring. They are both uncomfortable regarding their lack of interaction with others and sensitive to the judgments others make of them. Often they will not have a well-developed social support group to help them overcome

difficulties. Persons with moderate scores on 0 are dependable, conservative, cautious, unoriginal, serious, and overcontrolled. Normal males who score high on 0 are described as modest, inhibited, lacking in self-confidence, and generally deficient in poise and social presence. Normal females who score moderately high are somewhat similarly described as modest, shy, self-effacing, sensitive, and prone to worry.

Treatment Implications The relative elevation on 0 is potentially quite useful in treatment planning because it provides an index of the individual's degree of social comfort, inhibition, and control in relationships. It also indicates the degree to which the person is able to become engaged in interpersonal relationships. Thus, high scorers would have difficulty engaging in therapy because they are shy, withdrawn and anxious, whereas low scorers would have difficulties as well because of a superficial orientation and disinclination to reflect inwardly. High scorers would take time to develop a therapeutic relationship and expect the therapist to be directive and dominate. A therapist who is somewhat withdrawn and nondirective might increase such a client's anxiety and the person might then discontinue therapy. On the surface, these clients might appear to be unmotivated and passive but internally they are likely to feel high-strung and anxious (check LSE/Low Self-Esteem, A-lse/Low Self Esteem, SOD/Social Discomfort, and A-sod/Social Discomfort content scales). They are also likely to be overcontrolled and experience considerable difficulties in making changes. Group treatment and social skills training are often appropriate interventions. However, the group should be supportive and accepting, thereby increasing the likelihood that they would experiment with new behaviors.

MMPI-A Considerations Among adolescents, high scores on 0 are a clear indication of difficulties in social relationships, particularly related to low self-esteem and social withdrawal. The behavioral correlates for girls suggest that they are withdrawn, shy, fearful, depressed, may have had suicidal ideation and/or gestures, have eating problems, are socially withdrawn, and have only a few friends. Elevations also suggest an inhibitory effect in that they are unlikely to actually act out on their pathology. Thus, they rarely report difficulties with drugs or alcohol, delinquency, sexual acting out, and have little interest in heterosexual relationships. There are less behavioral correlates for boys, but high scores do suggest that they are unlikely to participate in school activities.

Low Scores on Scale 0

Low scorers are described as warm, outgoing, assertive, self-confident, verbally fluent, and gregarious. They will have strong needs to be around other people. They are likely to be concerned with power, recognition, and status. They may even be opportunistic, exhibitionistic, manipulative, and self-indulgent. In some cases, they might be immature, self-indulgent and superficial in their relationships with others. Normal males who score low are often perceived as being sociable, expressive, socially competitive, and verbally fluent. Normal females are described similarly as sociable, talkative, assertive, enthusiastic, and adventurous. Extremely low scores suggest persons who have highly developed social techniques; but behind their external image, they

may have feelings of insecurity with a strong need for social approval. They may also be hypersensitive and may have difficulties dealing with feelings of dependency. They are likely to have a large number of superficial friends, but probably do not feel close to anyone.

TWO-POINT CODES

Code-type interpretation often produces more accurate and clinically useful interpretations than merely interpreting individual scales. The basis of code-type interpretation depends on empirical correlations among various classes of nontest behavior. The two-point codes included in the following section have been selected based on their frequency of occurrence, the thoroughness of the research performed on them, and their relative clinical importance. Thus, some combinations of code types will not be discussed.

Code-type interpretation is most appropriate for disturbed populations in which *T* score elevations are at least 65 on the MMPI-2 or MMPI-A. The descriptions are clearly oriented around the pathological dimensions of an individual. The two-point code descriptions, then, do not have the same divisions into low, moderate, and high elevations as the individual scores but are directed primarily toward discussions of high elevations. When considering two-point codes that are in the moderate range (MMPI-2 $T = 60–70$), interpretations should be made with caution and the more extreme descriptions should be considerably modified or excluded.

Usually, the elevation of one scale in relationship to the other does not make much difference as long as the elevations are still somewhat similar in magnitude. A general approach is that, if one scale is 10 points or more higher than the other, the higher one gives more color to, or provides more emphasis for, the interpretation. Specific elaborations are made for scales in which a significant difference between their relative elevations is especially important. If the scales have an equal magnitude, they should be given equal emphasis.

In some cases, more than two scales will be equally elevated, thereby making it difficult to clearly establish which scales represent the two-point code. In these cases, clinicians should look at the descriptions provided for other possible combinations. For example, if scales 2, 7, and 8 are elevated for a particular profile, the clinician should look up the 27/72 code as well as codes 78/87 and 28/82. The descriptions for the code type with the highest elevations and those descriptors that are common between the different code descriptions are most likely to be valid. However, multiple elevations also raise the issue of the generalizability of the MMPI descriptors (which the majority of research has been derived from) and the MMPI-2 (Butcher et al., 1990; Butcher & Williams, 1992; D. Edwards et al., 1993; Humphrey & Dahlstrom, 1995; Tellegen & Ben-Porath, 1993). Up to 50% of the code types have been found to differ, which is particularly true for poorly defined code types. This would potentially compromise the validity of the code type descriptions. A more cautious approach would be to rely more on the single-scale descriptors.

In developing meaningful interpretations, it is important to continually consider the underlying significance of elevated scales. This requires considering factors such as the

manner in which the scales interact, the particular category of psychopathology they suggest, and their recurring patterns or themes. When possible, *DSM-IV* classifications have been used, but the term *neurosis* is used occasionally because of its ability to summarize a wide variety of disorders and/or its ability to refer to a cluster of related scales (e.g., "neurotic triad"). Some characteristics described in the code types will be highly accurate for a specific person, whereas others will not be particularly relevant or accurate. Clinicians, then, will need to continually reflect on their data to develop descriptions and diagnoses that are both accurate and relevant.

Use of the code types from the MMPI-A should be made with considerable caution because there is currently insufficient research on the behavioral correlates of these code types. In contrast, there is considerable research on the correlates of individual MMPI-A scale elevations. With this caution in mind, it is recommended that clinicians tentatively use the code types described in the following pages to help generate hypotheses concerning adolescent functioning. This is partially justified in that many of the MMPI code type correlates are common for both adults and adolescents (Archer, 1992a). In addition, the majority of the code types derived from the MMPI will also be the same for the MMPI-A, especially if these code types are well defined. If there are differences between adult and adolescent descriptors, or if no adolescent descriptors are available, this will be noted in the following code type descriptions.

12/21

Symptoms and Behaviors

Difficulties experienced by patients with the 12/21 code type revolve around physical symptoms and complaints that can be either organic or functional (check the HEA/Health Concerns content scale). Common complaints relate to pain, irritability, anxiety, physical tension, fatigue, and overconcern with physical functions. In addition to these symptoms, a significant level of depression is present. These individuals characteristically handle psychological conflict through repression and by attending to real, exaggerated, or imagined physical difficulties. Regardless of whether these physical difficulties are organically based, these individuals will exaggerate their symptoms and use them to manipulate others. In other words, they elaborate their complaints beyond what can be physically confirmed, often doing so by misinterpreting normal bodily functions. Typically, they have learned to live with their complaints and use them to achieve their own needs. This code pattern is more frequently encountered in males and older persons.

The three categories of patients that this code is likely to suggest are the generalized hypochondriacs, the chronic pain patients, and persons having recent and severe accidents. General hypochondriacs are likely to have significant depressive features and to be self-critical, indirect, and manipulative. If their difficulties are solely functional, they are more likely to be shy and withdrawn, whereas persons with a significant organic component are likely to be loud complainers. Furthermore, complaints are usually focused around the trunk of the body and involve the viscera. This is in contrast to the 13/31 code, in which complaints are more likely to involve the central nervous system and peripheral limbs. When the 12/21 code is produced by chronic pain patients with an organic basis, they are likely to have given in to their pain and learned to

live with it. Their experience and/or expression of this pain is likely to be exaggerated, and they use it to manipulate others. They may have a history of drug or alcohol abuse, which represents attempts at "self-medication." The most common profile associated with heavy drinkers consists of elevations in Scales 1, 2, 3, and 4. Such persons will experience considerable physical discomfort, digestive difficulties, tension, depression, and hostility, and will usually have poor work and relationship histories. The third category associated with the 12/21 code type involves persons who are responding to recent, severe accidents. Their elevations on Scales 1 and 2 reflect an acute, reactive depression that occurs in response to the limiting effects of their condition.

Personality and Interpersonal Characteristics

The 12/21 clients are typically described as introverted, shy, self-conscious, and passive dependent. They may harbor resentment against persons for not providing them with sufficient attention and emotional support. Interpersonally, they are likely to be extremely sensitive and manipulate others through references to their symptoms.

Treatment Implications

The 12/21 clients lack insight, are not psychologically sophisticated, and resent any implications that their difficulties may be even partially psychological (check the TRT/Negative Treatment Indicators scale). It is difficult for them to take responsibility for their behavior. They somatize stress, and one result is that they are able to tolerate high levels of discomfort before being motivated to change. Thus, they are not good candidates for psychotherapy, especially if the therapy is insight oriented. Typically, they seek medical explanations and solutions for their difficulties.

13/31

Symptoms and Behaviors

The 13/31 code type is associated with the classic conversion V, which occurs when Scale 2 is significantly lower (10 points or more) than Scales 1 or 3. As 2 becomes lower in relation to 1 and 3, the likelihood of a conversion disorder increases. This type of difficulty is strengthened in males who have correspondingly high Scales 4 and 5, and in females with a correspondingly high 4 but lowered 5. However, the 13/31 code type is more frequent in females and the elderly than in males and younger persons. Typically, very little anxiety is experienced by individuals with these profiles because they are converting psychological conflict into physical complaints. This can be checked by looking at the corresponding elevations of Scales 2 and 7. If these are also high, they are experiencing anxiety and depression, perhaps because their conversions are currently unable to effectively eliminate their conflicts. Persons with conversion Vs will typically engage in extensive complaining about physical difficulties. Complaints may involve problems related to eating, such as obesity, nausea, anorexia nervosa, or bulimia; and there may be the presence of vague "neurological" difficulties, such as dizziness, numbness, weakness, and fatigue. There is often a sense of indifference and a marked lack of concern regarding these symptoms. These individuals have a strong need to appear rational and socially acceptable, yet nonetheless control others

through histrionic and symptom-related means. They defensively attempt to appear hypernormal, which is particularly pronounced if the *K* scale is also elevated. Regardless of the actual, original cause of the complaints, a strong need exists to exaggerate them. Even if their complaints were originally caused by an organic impairment, there will be a strong functional basis to their problems.

If Scale 3 is higher than Scale 1, this allows for the expression of a certain degree of optimism, and their complaints will most likely be to the trunk of the body. Thus, patients might complain of difficulties such as gastrointestinal disorders, or diseases of the lungs or heart. Furthermore, a relatively higher 3 suggests the strong use of denial and repression. These people are passive, sociable, and dependent; they manipulate others through complaints about their "medical" problems. Conversely, if Scale 3 is lower than Scale 1, the person tends to be significantly more negative, and any conversion is likely to be to the body extremities such as the hands or legs. If scores are very high on Scale 8, a corresponding peak on Scale 1 is associated with somatic delusions.

Under stress, their symptom-related complaints will usually increase. However, when the stress level decreases, their symptoms will often disappear.

The most frequent diagnoses with 13/31 codes are major affective disorders (major depression, dysthymic disorder) hypochondriasis, conversion disorder, passive-aggressive personality, and histrionic personality. Anxiety may be present if either Scale 7 or 8 is also elevated, but these corresponding elevations are rare. The 13/31 profile is also found in pain patients with organic injuries, whose symptoms typically worsen under stress. Malingering of somatic complaints might be indicated if potential gain is a factor and 13/31 is quite high (especially if 3 is above $T = 80$) even if *F* is not elevated (because they want to emphasize their psychological normality but exaggerate the specifically physical nature of their difficulties).

Personality and Interpersonal Characteristics

Interpersonal relationships will be superficial, with extensive repression of hostility, and often their interactions will have an exhibitionistic flavor. Others describe them as selfish, immature, and egocentric but also as being outgoing, extraverted, and with strong needs for affection. They typically lack insight into their problems, use denial, and will often blame others for their difficulties (check the Repression/*R* scale). Usually, they are extremely threatened by any hint that they are unconventional and tend to organize themselves around ideals of service to others. However, their relationships and actual degree of involvement tend to be superficial. They may also feel resentment and hostility toward persons they feel have not provided them with sufficient attention and emotional support. When the conversion V is in the normal range (1 and 3 at or slightly below 65 on the MMPI-2), persons will be optimistic but somewhat immature and tangential. They can be described as responsible, helpful, normal, and sympathetic.

Treatment Implications

Because they lack insight and need to appear hypernormal, they typically make poor candidates for psychotherapy. They prefer simple, concrete answers to their difficulties and avoid introspection. However, they might respond to either direct suggestions or placebos, especially if the placebos are given in a medical context. Interventions

such as stress inoculation to reduce their stress might also be helpful. A potentially useful technique is to describe any psychosocial interventions using medical terminology. Thus, biofeedback or other stress reduction techniques might be referred to as *neurological retraining*. Often, however, they will terminate treatment prematurely, especially if their defenses are challenged. This issue becomes all the more difficult if there is a personality disorder as this would require a lengthier commitment to therapy.

14/41

Symptoms and Behaviors

The 14/41 code is encountered somewhat rarely, but is important because persons with these elevations will be severely hypochondriacal. They will be egocentric, will demand attention, and will express continuous concern with their physical complaints. There will be some similarities to other high-scoring 4s in that these individuals may have a history of alcohol abuse, drug addiction, and poor work and personal relationships (check WRK/Work Interference, MAC-R/MacAndrew Alcoholism scale, APS/Addiction Potential scale, and AAS/Addiction Acknowledgment scale) to refine interpretations). They may also be indecisive and rebellious.

The two most frequently encountered diagnoses will be hypochondriasis and a personality disorder, especially antisocial personality. Differentiation between these two can be aided by noting the relative strength of either Scale 1 or 4, as well as other related scales. Profiles involving "neurotic" features (anxiety, somatoform, dissociative, and dysthymic disorders) are characterized by a relatively higher Scale 1 with 2 and/or 3 also elevated. Personality disorders are more strongly suggested when Scale 4 is the primary high point.

Personality and Interpersonal Characteristics

A core feature of this code type is likely to be ongoing personality difficulties. They are likely to act out and use poor judgment. Interpersonal interactions will be extremely manipulative but rarely will they be extremely antisocial. They might feel a sense of rebelliousness toward their homes and parents although these feelings are not likely to be expressed openly. Although they will be able to maintain control over their impulses, they will do so in a way that is bitter, pessimistic, self-pitying, and resentful of any rules and limits that are imposed on them. They are likely to be described by others as demanding, grouchy, and dissatisfied (check the CYN/Cynicism, ASP/Antisocial Practices, and FAM/Family Problems scales).

Treatment Implications.

Usually, 14/41 patients will be resistant to therapy, although they may have a satisfactory response to short-term, symptom-oriented treatment. However, long-term therapy will be difficult and characterized by sporadic participation. Sessions may become somewhat tense because of their level of resentment and hostility, which is likely to be sometimes expressed toward the therapist (check the TRT/Negative Treatment Indicators and ANG/Anger scales).

18/81

Symptoms and Behaviors

Peaks on Scales 1 and 8 are found with persons who present a variety of vague and unusual complaints (check the HEA/Health Concerns scale). They may also experience confusion, disorientation, and difficulty in concentrating. They focus on physical symptoms as a way to organize their thoughts, although the beliefs related to these symptoms may represent delusions. Their ability to deal effectively with stress and anxiety is extremely limited. They will experience interpersonal relationships with a considerable degree of distance and alienation. Often, they will feel hostile and aggressive but will keep these feelings inside. However, when such feelings are expressed, the expressions will be made in an extremely inappropriate, abrasive, and belligerent manner. Others will perceive these individuals as eccentric or even bizarre. They will distrust others and may disrupt their relationships because of the difficulty in controlling their hostility. There may even be paranoid ideation, which will probably, but not necessarily, be reflected in an elevated Scale 6. They might be confused, distractible, and disoriented.

Common scales that are elevated along with 1 and 8 are 2, 3, and/or 7. These serve to color or give additional meaning to 18/81. Thus, an elevated Scale 2 will emphasize self-critical, pessimistic dimensions; 7, the presence of fears and anxiety (check the ANX/Anxiety, A/Anxiety, FRS/Fears, and OBS/Obsessions scales); and 3, the likelihood of conversions and/or somatic delusions.

The 18/81 code is frequently diagnosed as schizophrenia, especially if the *F* scale is also high. With a normal *F,* hypochondriasis is an important possibility, but if Scale 7 is elevated, an anxiety disorder is also strongly suggested.

Personality and Interpersonal Characteristics

Personality difficulties of a long-standing nature are likely to be a significant factor. The 18/81 clients are low in interpersonal trust and feel socially inadequate. They will feel socially isolated and alienated. Consistent with this, their histories will often reveal a nomadic lifestyle with poor work histories (check the WRK/Work Interference scale).

Treatment Implications

Engaging them in therapy will be difficult because their level of insight will be poor. In addition, they will be distrustful, pessimistic, alienated, and hostile (check the TRT/Negative Treatment Indicators scale).

19/91

Symptoms and Behaviors

The 19/91 code is rarely encountered but is important because it may suggest organic difficulties relating to endocrine dysfunction or the central nervous system. Complaints are likely to include gastrointestinal difficulties, exhaustion, and headaches. There will be extensive complaining and overconcern with physical difficulties, but these patients may paradoxically attempt to deny and conceal their complaints at the same time. In other words, they may invest significant energy in avoiding confrontations relating to

their complaints, yet will make a display of these techniques of avoidance. They will typically be extraverted, talkative, and outgoing, but also tense and restless. They might be in a state of turmoil and experience anxiety and distress. The expectations they have of themselves will be extremely high, yet their goals will be poorly defined and often unobtainable. If their complaints have no organic basis, their behavior may be an attempt to stave off an impending depression. Often, this depression will be related to strong but unacceptable dependency needs.

Both hypochondriasis and manic states are frequent diagnoses and may occur simultaneously. These may be in response to, and exacerbated by, an underlying organic condition, an impending depression, or both. Corresponding elevations on Scales 4 and 6 make the possibility of a passive-aggressive personality an important diagnostic consideration.

Personality and Interpersonal Characteristics

Superficially, these clients might appear outgoing, assertive, and ambitious. However, they are likely to have an underlying passive dependent core to their personalities.

Treatment Implications

Psychotherapy will be difficult because these individuals are reluctant to accept a psychological explanation for their complaints (check the TRT/Negative Treatment Indicators scale).

23/32

Symptoms and Behaviors

Persons with elevations on Scales 2 and 3 will be lacking in energy, weak, apathetic, listless, depressed, anxious, and frequently report gastrointestinal complaints. They feel inadequate and have difficulty accomplishing their daily activities. Much of their energy is invested in excessively controlling their feelings and behavior. Although situational stress may serve to increase their depression, usually this depression is longstanding, and they have learned to live with their unhappiness and general lack of satisfaction. They are not very involved or interested in life and experience a difficult time initiating activities.

Some important male-female differences exist in the expression of this code type. Males are more ambitious, industrious, serious, and competitive, but also are immature and dependent. They strive for increased responsibilities, yet also fear them. They want to appear normal and receive recognition for their accomplishments, yet they often feel ignored and their level of work adjustment is often inadequate. In contrast, females are more apathetic and weak, and experience significant levels of depression. They have usually resigned themselves to long-term unhappiness and a lack of satisfaction. Although there is often significant marital strife (check the FAM/Family Problems scale), they rarely seek divorce.

Affective disorders represent the most frequent category of diagnosis given to this code. Corresponding elevations on Scales 4, 6, and 0 may provide additional information relating to the personality of these persons. With a high Scale 4, there is more

likely to be an angry, brooding component to their depression, with underlying antisocial thoughts, yet their external behavior is usually overcontrolled. An elevated Scale 6 suggests that their depression relates to extreme interpersonal sensitivity and distrust, whereas a high 0 indicates they are socially withdrawn and introspective. An additional diagnosis that should be considered is a major depression with psychotic features, especially if Scales *F* and/or 8 are also elevated. Many patients with this code type are diagnosed as having a somatoform disorder. A 23/32 code type is frequently seen with chronic pain patients, especially if Scale 1 is also elevated.

Personality and Interpersonal Characteristics

Individuals having this code type are often perceived as passive, docile, and dependent; therefore, they often obtain nurturance from others. By keeping their relationships superficial, they achieve a certain level of security. Their behavior often elicits nurturance from others. They are uncomfortable around members of the opposite sex and may experience sexual maladjustment, including impotence or frigidity. Interpersonally, they appear immature, childish, and socially inadequate. In terms of work, they feel the need to achieve and be successful, but are afraid of the added pressure this might produce. Although they might appear as if they are driven to succeed, they are anxious regarding competitive situations. Despite this avoidance of competition, they feel that their achievements are not adequately recognized.

Treatment Implications

These individuals will rarely volunteer for psychotherapy, their level of insight is poor, and they usually do not show significant improvement during treatment. This is primarily because their main dynamic is denial and situations such as therapy represent a threat to their avoidant style. Any conflicts are likely to be somatized, and they are highly invested in medical explanations for their complaints. Accordingly, they might seek medical "solutions" to interpersonal conflicts through methods such as tranquilizers and pain medications. A further area that makes treatment difficult is their inability to tolerate a considerable amount of discomfort and seem resigned to live with their unhappiness. However, because their level of distress is usually quite high, some method of symptom relief is indicated, possibly through antidepressant medication. In addition, supportive (rather than insight-oriented) psychotherapy is often beneficial.

24/42

Symptoms and Behaviors

The most significant aspect of the 24/42 code is the underlying antisocial trend to the clients' personalities, with difficulty maintaining control over their impulses. However, when they act on their underlying antisocial impulses, they experience guilt and anxiety regarding the consequences of their actions. This anxiety usually occurs too late to serve as an effective deterrent, and these individuals are unable to plan ahead effectively. The depression they experience is probably situational, and the distress they feel may reflect a fear of external consequences rather than an actual internalized moral code. When the situation has subsided, there is usually further acting out. For

this reason, the 24/42 code is sometimes considered to reflect an antisocial personality who has been caught.

The history of persons with high Scales 2 and 4 is often characterized by heavy drinking and/or drug abuse, which serves as a form of self-medication for their depression (check the MAC-R, ACK/Alcohol Acknowledgment, and APS/Alcohol Potential scales). Their interpersonal relationships are poor, which is reflected in numerous family difficulties (check the FAM/Family Problems scale) and sporadic employment. Their prospects for long-term employment are rarely favorable (check the WRK/Work Interference scale). These problems have often resulted in numerous legal complications (check the ASP/Antisocial Practices scale).

The hostility that is present with the 24/42 code may be expressed either directly or indirectly. A more direct expression is suggested if Scale 6 is high, as these individuals may feel justified in externalizing their anger because of real or imagined wrongs that have been committed against them. In contrast, a low 6 may reflect a suppression or unconscious denial of hostility. If high energy levels are suggested by a high Scale 9, the person may be extremely dangerous and volatile, and may have committed violent behaviors.

The 24/42 code is associated with personality disorders, especially passive-aggressive or antisocial personalities. This is further strengthened if Scale 6 is also high. However, this code frequently reflects an adjustment disorder with a depressed mood. An important distinction to make is whether the depression is reactive or chronic. If chronic, difficulties related to anxiety, conversions, and depression (neurotic features) will more likely be predominant, especially if Scales 1 and 3 are also high. A reactive depression is more likely to represent an antisocial personality who has been apprehended for his or her impulsive acting out. Substance abuse may be either the primary difficulty or may occur in addition to the other disorders suggested earlier. If Scale 4 is extremely elevated (above 90), a psychotic or prepsychotic process may be present, especially if *F* and 8 are also high.

Personality and Interpersonal Characteristics

The initial impression may be friendliness or even charm, and, in a hospital setting, these patients may attempt to manipulate the staff. At their best, they can appear sociable, competent, and enthusiastic. Others might perceive them as sociable and outgoing. However, in the long term, they are likely to produce resentment in interpersonal relationships. While appearing superficially competent and confident, they are likely to experience an underlying sense of dissatisfaction and feel self-conscious. Such persons respond to their failures with pessimism, self-criticism, and self-doubt. In an attempt to deal with these feelings, they will often develop passive-dependent relationships.

Treatment Implications

A 24/42 code type is the most frequent pattern found in alcohol and drug treatment programs. As a result, persons with this code type should always be assessed for substance abuse, regardless of the setting or reason for referral. Often an acknowledgment by clients that they indeed do have a drug or alcohol problem and an appraisal into its impact on their lives are essential initial steps (check the AAS/Addiction Acknowledgment scale). This profile also suggests long-standing personality difficulties that often

make therapy difficult. Although such people may promise to change and their guilt is generally authentic, their acting out is usually resistant to change. Effective therapy must include clear limits, a change in environment, warm supports, and continual contact. However, the prognosis for long-term success in therapy is poor and the individuals will be likely to terminate when confronted with situational stress or when external motivators (e.g., legal) have been eliminated. Thus, some sort of external monitoring (i.e., legal or work-related) of their treatment, perhaps even conducting their treatment in a controlled environment, is advisable. Because peer influences are likely to have considerable impact, group interventions are likely to be more effective than individual treatment.

26/62

Symptoms and Behaviors

The most significant feature of the 26/62 code is extreme sensitivity to real or imagined criticism. These individuals will sometimes interpret the statements of others in a way that creates rejection, yet their conclusions will be based on insufficient data. Even minor criticism is brooded over and elaborated on. Usually, they have long histories of difficulties with interpersonal relationships. Others describe them as resentful, aggressive, and hostile. To protect themselves from the impending rejection of others, they will often reject others first, which results in other people avoiding them. When they are avoided, these individuals then have evidence that they are being rejected, which gives them a justification for feeling and expressing anger. They can then blame others for their difficulties. This cycle is thus self-fulfilling and self-perpetuating, yet such people have difficulty understanding the part they play in creating the interpersonal responses directed toward them.

If Scales 7, 8, and possibly 9 are also high, a greater likelihood of a psychotic or prepsychotic condition exists, especially paranoid schizophrenia. A more controlled, well-defined paranoid system with a generally adequate level of adjustment may be suggested when Scales 2, 6, and *F* are only moderately elevated. Further possible diagnoses with the 26/62 code are a dysthymic disorder and, if Scale 4 is also elevated, a passive-aggressive personality.

Personality and Interpersonal Characteristics

Because persons with this code type are openly hostile and hypersensitive, they are likely to have poor interpersonal relationships (check the FAM/Family Problems and CYN/Cynicism scales). They are blaming, resentful, hostile, and are likely to have passive-aggressive qualities. These patterns are usually of a long-standing nature and are difficult to alter.

Treatment Implications

The major challenge will be to effectively develop and maintain their rapport and trust. This will mean continually disengaging from their hostility and suspiciousness (check the ANG/Anger scale). An important area of further assessment is to determine the extent of possible underlying psychotic processes.

27/72

Symptoms and Behaviors

The 27/72 code is extremely common in psychiatric populations and reflects persons who are depressed, agitated, restless, and nervous. Their behavior may be accompanied by slowed speech and movements, as well as by insomnia and feelings of social and sexual inadequacy. They generally spend a good deal of time anticipating problems before they actually occur and are vulnerable to actual or imagined threats. They worry excessively, often overreacting to minor events. Scales 2 and 7 reflect the relative degree of subjective turmoil the person is experiencing and they are thus often referred to as the *distress scales* (check the ANX/Anxiety, A/Anxiety, FRS/Fears, and OBS/Obsessiveness scales). Physical complaints may include weakness, fatigue, chest pain, constipation, and dizziness (check the HEA/Health Concerns scale).

Moderate elevations on Scales 2 and 7 can indicate a good prognosis for therapy, because this suggests that the person is introspective and is experiencing a sufficient amount of distress to be motivated to change. However, extreme elevations are likely to reflect a high level of disruption in his or her ability to cope. The most frequent diagnoses are affective disorders, particularly major affective disorder, although they might also have an adjustment disorder with depressed mood. Anxiety disorders are also a possibility, particularly obsessive-compulsive disorder. Possible personality disorders might be avoidant, compulsive, or passive-aggressive. However, with only moderate elevations, they may be normals who are fatigued and exhausted, with a high degree of rigidity and excessive worry. This code occurs more frequently with males 27 years or older from higher educational backgrounds. If 4 is elevated along with 2 and 7 (274/427/724), the meaning of the profile is changed. It then suggests persons who are anxious and depressed because of poor judgment related to self-indulgence, particularly related to problem alcohol or drug use (check the MAC-R, AAS/Addiction Acknowledgment, and the APS/Addiction Potential scales).

Personality Characteristics

These clients can be characterized as perfectionistic and meticulous, and as having a high need for recognition. Their thinking is often obsessive, and they experience a wide variety of phobias and fears (check the FRS/Fears scale). Interpersonally, they have difficulty asserting themselves and will be self-blaming, self-punishing, and passive-dependent (check the SOD/Social Discomfort scale). They will rarely be argumentative or provocative. Their consciences are strong and inflexible, and they will often be extremely religious in a rigidly fundamental manner. Most are married and their courtships were fairly brief, many marrying within one month of their initial dating. They are described by others as docile and dependent, and typically elicit nurturance from others. They frequently rely on their friends and family to an excessive extent. Internally, they feel inadequate, insecure, and deal with feelings of hostility in an intropunitive manner.

Treatment Implications

Although 27/72 persons usually express a great deal of pessimism regarding treatment and the future in general, their psychological distress is ordinarily reactive, and in

time, they can be expected to improve. For most patients having this profile, the disorder takes between one month and one year to develop, and, if they report for treatment, it will be their first need for such intervention. If these scales are extremely high, the person may be too agitated to focus and concentrate. In such cases, medication may be necessary to relax him or her sufficiently to function in a psychotherapeutic context. The presence of suicidal thoughts is a definite possibility, especially if Scales 6 and 8 are also elevated, and the suicidal potential of these patients must be carefully evaluated. They can often be extremely self-critical during therapeutic sessions and require considerable emotional support. They are prone to being perfectionistic and guilty, which frequently leads to unproductive periods of rumination. While obsessive about the possibility of change, they often have a difficult time actually attempting new behaviors. However, they generally establish new relationships relatively easily, and these relationships are frequently deep and of a long duration.

When working with persons with 274/427/724 code types, their drinking patterns might be of a long-standing nature, therefore complicating any interventions. The possible presence of these difficulties should be determined early in the treatment sessions. In contrast to the pure 27/72 code type, they do not do well in individual insight-oriented therapy and are likely to terminate prematurely. There may be an initial "honeymoon" effect in which changes have apparently been made, but during times of stress, they are likely to act out and undermine any progress. They would be most likely to benefit from group interventions with a focus on clear, specific goals that would include, among other things, environmental changes.

28/82

Symptoms and Behaviors

Persons with the 28/82 code complain of depression, anxiety, insomnia, fatigue, and weakness, as well as mental confusion, memory impairments, and difficulties in concentrating. They may also feel withdrawn, alienated, agitated, tense, and jumpy. Their motivation to achieve is characteristically low, as is their overall level of efficiency. They are likely to be unoriginal, stereotyped, apathetic, and indifferent. Often, they will have fears relating to an inability to control their impulses, including suicide. They are suspicious and extremely sensitive to the criticisms of others. Delusions and hallucinations may also be present, especially if Scale 8 is greater than 85. This list of complaints presents a highly diverse description of attributes, only some of which may be present in any specific case. The presence or absence of these complaints must be determined by examining data other than mere scale elevations. This may include the investigation of critical items, clinical interview data, personal history, and the use of the Harris-Lingoes and content scales (particularly BIZ/Bizarre Mentation, FRS/Fears, OBS/Obsessions, LSE/Low Self-Esteem, and SOD/Social Discomfort).

Differential diagnosis can be extremely important to determine. Most persons with this code type are diagnosed as having a major affective disorder (bipolar-depressed or major depression). Schizophrenia or schizoaffective disorder is also a possibility. Personality disorders might include borderline, avoidant, obsessive-compulsive, or schizoid. These personality patterns might feature lability, emotional instability, and acting out.

Personality and Interpersonal Characteristics

Relevant personality descriptors include resentful, unassertive, dependent, and irritable. They often feel excessive guilt and are self-punitive. They justifiably have a fear of losing control of their emotions. A typical coping strategy is to deny unacceptable impulses, but this sometimes results in dissociative periods of acting out.

Treatment Implications

These clients are likely to have multiple problems related to expressing their anger, relationship difficulties, and social withdrawal. In particular, they might lose control over their feelings of anger, which might be directed toward the therapist during times of stress. They are also likely to feel ambivalence toward relationships in general, and this may express itself in resistance to therapy. This ambivalence will also make it difficult to experiment with new strategies learned in therapy. Thus, therapy tends to be long term. The therapist potentially can provide a point of stability in an otherwise chaotic and unpredictable life. An important area that should be assessed both during the initial session(s) and throughout treatment is the potential for suicide. During times of crises, many persons with this profile might require medication to control their thoughts and feelings.

29/92

Symptoms and Behaviors

Although anxiety and depression are present with the 29/92 code, a high level of energy also predominates. This energy may be associated with a loss of control, or it may also serve to defend against experiencing underlying depressive feelings. By speeding up their level of activity, these individuals can distract themselves from unpleasant depressive experiences. At times, this will be successful, but they may also use alcohol either to relax or to decrease their depression. With moderate elevations, this code will, at the very least, reflect tension and restlessness. Often, these persons will ruminate on feelings of worthlessness. They are typically perceived as self-absorbed and self-centered. Somatic complaints (especially upper-gastrointestinal) and sporadic alcohol abuse are common. They have high needs for achievement but may paradoxically set themselves up for failure. When this code type occurs among younger persons, it might reflect a vocational crisis with a resulting loss of identity. Sometimes brain-injured persons have this profile, which reflects their feeling of loss of control over their thoughts and feelings, but they attempt to compensate by speeding up their level of activity.

If both scales are in the higher elevations, a mixed bipolar depression is suggested. However, both scales can change according to the particular phase the patient is in. This code can also reflect certain types of brain-injured patients or a cyclothymic disorder.

Personality and Interpersonal Characteristics

The core feelings will be a sense of inadequacy and worthlessness. However, the person may deny these feelings and defend against them with excessive activity.

Treatment Implications

Because alternating periods of intense activity followed by exhaustion and depression often occur, a major challenge of treatment is to stabilize these mood and activity swings. This might be further complicated by a long-standing history of alcohol or drug abuse. In addition, suicide potential should be carefully monitored. During initial assessment, depression may not be immediately apparent. However, a careful consideration of the client's background will reveal long-term but sporadic phases of depression.

34/43

Symptoms and Behaviors

Persons having peaks on Scales 3 and 4 are immature and self-centered, with a high level of anger that they have difficulty expressing. Thus, their anger will often be expressed in an indirect, passive-aggressive style. Outwardly, such individuals are continually trying to conform and please other people, but they still experience a considerable degree of anger and need to find ways of controlling or discharging it. This anger stems from a sense of alienation and rejection by family members. They might at times vicariously act out their aggression by developing a relationship with an individual who directly and spontaneously expresses his or her hostility. Such a relationship might be characterized by the 34/43 individual's covertly encouraging and fueling the other person's angry expressions, yet on a more superficial social level, disapproving of the other person. Typically, these individuals will have poor insight regarding their own behavior. If Scale 6 is also high, their lack of insight will be even more pronounced because their hostility will be projected onto others. Usually, past interpersonal relationships have been difficult. There may be a history of acting out, marital discord, and alcohol abuse (check the MAC-R, AAS/Addiction Acknowledgment Scale, APS/Addiction Potential Scale, and MDS/Marital Distress scales). Females are more likely than males to have vague physical complaints such as headaches, blackouts, and upper-gastrointestinal complaints. Despite such complaints, these females are generally free from extensive levels of anxiety. Furthermore, their relationships will be superficial and will be characterized by naive expectations and a perfectionistic view of the world, which they maintain by glossing over and denying conflicts.

The 34/43 code most clearly fits the pattern of a passive-aggressive interactional style. However, histrionic or borderline personalities are also common. Persons with 34/43 code types are also frequently diagnosed as having an adjustment disorder with depressed mood or mixed emotional features. If both scales are extremely elevated (*T* greater than 85), there may be fugue states in which aggressive and/or sexual impulses will be acted out.

Personality and Interpersonal Characteristics

Conflicts relating to dependence versus independence are significant as both of these needs are intense. These individuals tend to demand approval and affection from others. However, they will also have underlying feelings of anger that can easily become activated by criticism. Superficially, they might appear conforming but underneath they have strong feelings of rebelliousness.

Treatment Implications

Treatment sessions are likely to be stormy because these clients will treat the therapeutic relationship similar to other relationships. Central issues will be self-control and difficulty with taking responsibility for their behaviors. The major resistance to therapy will be that they project blame onto others and have low levels of insight regarding this coping style. Often, they terminate therapy out of anger and frustration. Sometimes internal motivation to seek therapy is lacking, and they have been forced into treatment through external pressures from their spouses, work, or the legal justice system. Because they are relatively more responsive to peer (versus authority) pressures, group therapy can be quite effective. It is often useful to arrange for some external monitoring and external motivation to keep them in treatment.

36/63

Symptoms and Behaviors

A 36/63 code type indicates that the person is extremely sensitive to criticism, and represses his or her hostile and aggressive feelings. These individuals are fearful, tense, and anxious, and may complain of physical difficulties such as headaches or stomach problems. Overtly, they might deny suspiciousness and competitiveness, and might even see the world in naively accepting, positive, and perfectionistic terms. They can quickly and easily develop comfortable, superficial relationships. However, as a relationship's depth and closeness increases, their underlying hostility, egocentricity, and even ruthlessness become more apparent.

If Scale 6 is higher than Scale 3 (by more than 5 points), these individuals will attempt to develop some sense of security in their lives by seeking power and prestige. If Scale 3 is higher than Scale 6 (by more than 5 points), their tendency to blame will be reduced, and such people will be more likely to deny any conflicts or problems. This will be consistent with a tendency to idealize both themselves and their world. They will be more likely to develop somatic complaints rather than paranoid ideation, and the chance of a psychotic process is significantly reduced.

Personality and Interpersonal Characteristics

They will harbor feelings of resentment and hostility, especially toward family members although they are unlikely to express these feelings directly. At times, they can be naive and gullible.

Treatment Implications

Their ability to acquire personal insight is limited because they are psychologically unsophisticated and resent suggestions that their difficulties may be even partially psychological (check the TRT/Negative Treatment Indicators scale). They will usually blame their personal problems on others, which creates one of their major difficulties in relationships. In therapy, they will typically terminate abruptly and unexpectedly. They can be ruthless, defensive, and uncooperative. A central issue will be having them take responsibility for their feelings and behaviors.

38/83

Symptoms and Behaviors

This somewhat rare code involves symptoms of anxiety, depression, and complaints such as headaches, gastrointestinal disturbances, and numbness. They may have a series of obscure, intractable somatic complaints. If Scale 8 is significantly higher than Scale 3, these individuals may also have thought disturbances including mental confusion, disorientation, difficulties with memory, and, at times, delusional thinking (check the BIZ/Bizarre Mentation scale). They often experience considerable turmoil and feel tense, fearful, and worried. Outwardly, they might appear apathetic and withdrawn. Although they have unusual experiences related to their thought processes and feel socially alienated, they also have strong needs to appear normal and strong needs for affection. They feel that, if others knew how unusual their experiences were, they would be rejected. Thus, they are extremely afraid of dependent relationships. To protect themselves, they use extensive denial, which makes their capacity for insight poor. They typically will describe their difficulties in a vague, guarded, and nonspecific manner.

An important variation from the 38/83 code occurs when elevated Scales 3 and 8 are also accompanied by elevations on *K,* with a low *F.* Persons with this profile are likely to be affiliative, inhibited, and overconventional, and to have an exaggerated need to be liked and approved of by others. Frequently, they maintain an unrealistic yet unassailable optimism. They emphasize harmony, perhaps even at the cost of sacrificing their own needs, attitudes, and beliefs. Furthermore, individuals who have high 3s with low *F* scores are extremely uncomfortable with anger and will avoid it at all costs. Typically, they will also avoid independent decision making and many other situations in which they must exert their power. Because they have an exaggerated sense of optimism and deny their personal conflicts, these individuals rarely appear in mental health clinics. It is almost as if any feelings of anger, tension, or defeat are intolerable. Such feelings seem to represent both a personal failure and, perhaps more importantly, a failure in their attempts at controlling their world by developing an overconventional, exaggeratedly optimistic, and inhibited stance.

When Scale 3 is relatively higher than Scale 8, and 8 and/or *F* is less than 70, somatoform or dissociative disorders are important considerations. If 8 and *F* are both highly elevated, the person might be schizophrenic.

Personality and Interpersonal Characteristics

Persons with this profile can be described as having strong needs for attention and affection and are also immature and dependent. On the surface, they might seem conventional, stereotyped, and unoriginal. Despite having a number of unusual internal experiences, they are uncomfortable with these processes and tend to limit them by being intropunitive.

Treatment Implications

Because they are typically apathetic and uninvolved in life activities, it is similarly difficult to engage them in therapy. Treatment is further complicated because their level of insight is low. Specifically, they place considerable effort into appearing normal despite

considerable unusual underlying processes. Thus, individual insight-oriented therapy is contraindicated. However, they may be responsive to a more supportive and directive approach.

45/54

Symptoms and Behaviors

High scores on Scales 4 and 5 reflect persons who have difficulty incorporating societal values. For the most part, they can control antisocial feelings, but they may have brief episodes of acting out associated with low frustration tolerance and underlying anger and resentment. Their usual coping style is through passive-aggressive means. Overt homosexuals who make obvious displays of their orientation may have this code, especially if Scales 4 and 5 are the only peaks in an otherwise normal profile. The 45/54 code should in no way be considered diagnostic of homosexuality but simply, at times, is consistent with a subgroup of persons who have this orientation. To obtain further information associated with this or any profile in which Scale 5 is a high point, it is extremely helpful to interpret the third-highest scale and give it the degree of importance usually associated with the second highest point. Thus, a profile in which 4, 5, and 6 are all high might be interpreted as though it were a 46/64 code type.

Some important differences exist between males and females who have this code. First, it occurs much more frequently among men. Males with this code type will be openly nonconformist, but if they are from higher educational levels, they will be more likely to direct their dissatisfaction into social causes and express organized dissent toward the mainstream culture. If 9 is correspondingly high, they will be dissatisfied with their culture, sensitive, and aware, but will also have the energy to attempt to create change. They are often psychologically sophisticated, and can communicate clearly and openly. In contrast, elevated Scales 4 and 9, accompanied by a low Scale 5, suggest a high probability of sexual acting out and the probable development of a "Don Juan" personality. These men are self-centered and have difficulty delaying their gratification. Behind their overt display of affection is an underlying current of hostility.

Females with the 45/54 code will be openly rebelling against the traditional feminine role. Often, this rebellion is motivated by an intense fear related to developing dependent relationships. A further alternative interpretation is that these women are merely involved in a subculture or occupation that emphasizes traditionally male-oriented activities.

Personality and Interpersonal Characteristics

Persons with this profile are immature, self-centered, and inner-directed, and are not only nonconformist but also likely to openly express this nonconformity in a challenging, confrontive manner. They may also have significant problems with sexual identity and experience sexual dysfunction. A further area of conflict revolves around ambivalence relating to strong but unrecognized dependency needs.

Treatment Implications

Although persons with this profile are guarded and defensive about revealing themselves, they are also capable of thinking clearly and have good insight into their behavior.

They rarely report for treatment because they typically are satisfied with themselves and their behavior. They usually do not report being emotionally distressed. When they do seek treatment, issues are likely to center on dominance and dependence. Significant change is unlikely because of the chronic, ingrained nature of their personality.

46/64

Symptoms and Behaviors

Persons with the 46/64 code type are hostile, brooding, distrustful, irritable, immature, self-centered, and usually unable to form close relationships. They have significant levels of social maladjustment often related to continually blaming others for their personal faults. This style of blaming prevents them from developing insight into their own feelings and behavior, because they are constantly focusing on the behavior of others rather than their own. They lack self-criticism, and are highly defensive and argumentative, especially if *L* and *K* are also high. Although they lack self-criticism, they are highly sensitive to real or imagined criticism from others, often inferring hostility or rejection when this was not actually intended. To avoid rejection and maintain a certain level of security, they become extremely adept at manipulating others. Often, they will have a history of drug addiction or alcohol abuse (check the MAC-R, AAS/Addiction Acknowledgment, and APS/Addiction Potential scales).

Frequent corresponding high points are on Scales 2, 3, and/or 8. Males with high 8s are often psychotic, especially paranoid schizophrenic or prepsychotic, but with 2 and/or 3 also elevated, the chances of a borderline condition are significantly increased. These men are likely to be angry and to have significant conflicts relating to their own denied, but strong, needs for dependency. They are likely to rebel against authority figures and may use suicidal threats to manipulate others. Females with a 46/64 code type may be psychotic or prepsychotic, but they are more often passive-aggressive personalities. If Scale 3 is also elevated, they will have intense needs for affection and will be egocentric and demanding. However, they will be resentful of the demands placed on them by others.

Personality and Interpersonal Characteristics

A core issue is often passive dependency. These individuals frequently have adjustment difficulties associated with their hostility, anger, mistrust, and a tendency to blame others. They tend to avoid deep involvement. People perceive them as sullen, argumentative, obnoxious, and resentful of authority (check the ANG/Anger scale).

Treatment Implications

Persons with this profile are generally suspicious and even antagonistic toward treatment. When they do appear for treatment, it is at the insistence of someone else. As a result, they are mistrustful, suspicious, and project the blame for any difficulties onto someone else. Treatment plans should be concrete, clear, realistic, and described in a way that doesn't arouse suspicion or antagonism. A therapeutic relationship is difficult to establish and, once established, is likely to be somewhat turbulent. The possibility of angry acting out should be carefully monitored.

47/74

Symptoms and Behaviors

Persons with high scores on Scales 4 and 7 experience guilt over their behavior, and are brooding and resentful. Although they are frequently insensitive to the feelings of others, they are intensely concerned with their own responses and feelings. They justify this insensitivity because they feel rejected or restricted by others. Their behavioral and interpersonal difficulties follow a predictable cycle in which they will alternately express anger and then feel guilty over their behavior. While they feel angry, they may have little control over their behavior, which results in impulsive acting out (check the ASP/Antisocial Practices and ANG/Anger scales). This is followed by a phase of excessive overcontrol accompanied by guilt, brooding, and self-pity (check the O-H/Over-Controlled Hostility scale). Frustrated by these feelings, they may then attempt to selfishly meet their needs through means such as alcohol abuse, promiscuity, or aggressive acting out. Thus, the cycle continues and is usually fairly resistant to change. This frequently leads to legal problems and to difficulties in their work and home relationships. Although they do feel genuine and even excessive guilt and remorse, their self-control is still inadequate and their acting out continues.

Diagnostically, the 47/74 type is most likely to be either an antisocial personality or an anxiety disorder. This profile is frequently seen in alcohol, drug (check the MAC-R, AAS/Alcohol Acknowledgment, or APS/Alcohol Potential scales), or other treatment settings to which individuals with impulsive-compulsive styles are referred (e.g., eating disorder programs for persons with bulimia).

Personality and Interpersonal Characteristics

Core difficulties relate to feelings of insecurity and ambivalence regarding dependency. Clients need frequent reassurances that they are worthy.

Treatment Implications

During the early stages of treatment, clients typically show remorse and express the need to change. This might seem sincere but as their guilt diminishes, they will again act out. Thus, therapists should be suspicious of early "easy" gains. Frequently, the person will respond to limit-setting with anxiety and resentfulness, often either testing the limits or completely ignoring them. The style of acting out followed by guilt is a chronic pattern, and therapeutic attempts to decrease anxiety may actually result in an increase in acting out because the control created by guilt and remorse might be diminished. These individuals may respond well to reassurance and support. However, long-term, fundamental change will be difficult to achieve.

48/84

Symptoms and Behaviors

Persons with the 48/84 code are strange, eccentric, emotionally distant, and have severe problems with adjustment. Their behavior is unpredictable and erratic, and may involve strange sexual obsessions and responses. Usually, there will be antisocial

behavior resulting in legal complications (check the ASP/Antisocial Practices scale). These individuals also lack empathy, and are nonconforming and impulsive. Sometimes, they will be members of strange religious cults or unusual political organizations. In their early family histories, they learned that relationships were dangerous because of constant confrontation with intense family conflicts. They were rejected and, as a result, felt alienated and hostile, sometimes attempting to compensate with counterrejection and other forms of retaliation. Their academic and later work performance has usually been erratic and characterized by underachievement. In interpersonal relationships, their judgment is generally poor and their style of communication is likely to be inadequate. Often, others feel as if they are missing important elements or significant connotations of what the 48/84 individual is saying, but they cannot figure out exactly what or why.

If *F* is elevated with a low Scale 2, these individuals are typically aggressive, cold, and punitive, and have a knack for inspiring guilt and anxiety in others. Often, they take on roles in which such behavior is socially sanctioned, for example, a rigid law enforcement officer, overzealous member of the clergy, or a strict school disciplinarian. Their behavior may range all the way from merely stern, punitive, and disapproving, to actual clinical sadism. Underneath these overt behaviors, they usually have a deep sense of alienation, vulnerability, and loneliness, which may give rise to feelings of anxiety and discomfort.

Criminal behavior occurs frequently in males with a 48/84 code type, especially when Scale 9 is also elevated. The crimes are likely to be bizarre, and often extremely violent, involving homicide and/or sexual assault. These behaviors are usually impulsive, poorly planned, without apparent reason, and generally self-defeating, eventually resulting in self-punishment. Females are less likely to act criminally, but their relationships will usually be primarily sexual and they will rarely become emotionally close. Often, they will form relationships with men who are significantly inferior to themselves and who could be described as losers.

The most likely diagnosis is a schizoid or paranoid personality. However, a psychotic reaction, often paranoid schizophrenia, is also common, especially with elevations on Scale 6.

Personality and Interpersonal Characteristics

Although these individuals have deep needs for attention and affection, they frequently set themselves up for rejection and failure. They have deep feelings of insecurity and a poor self-concept.

Treatment Implications

Because the client will be aloof and unconventional, it will be difficult to establish a therapeutic relationship. The sessions are likely to be chaotic with difficulty focusing on relevant areas. Thus, they may seem relatively unproductive. There will often be so many different problems to work on that it is difficult to know where to begin, and it easy to get sidetracked. Treatment may be further complicated by long-standing drug- and alcohol-related problems. Acting out may further complicate the picture. Because these clients are also likely to be mistrustful, they are likely to terminate prematurely.

49/94

Symptoms and Behaviors

Persons with 49/94 codes not only feel alienated and have antisocial tendencies, but also have the energy to act on these tendencies. They can be described as self-indulgent, sensation seeking, impulsive, oriented toward pleasure, irritable, extraverted, violent, manipulative, and energetic. They have poorly developed consciences, with a marked lack of concern for rules and conventions. Because they are free from anxiety, talkative, and charming, they can often make a good initial impression. However, their relationships are usually shallow because any sort of deeper contact with them brings out the more problematic sides of their personality. An investigation of history typically reveals extensive legal, family, and work-related difficulties (check the ASP/Antisocial Practices and WRK/Work Interference scales). The 49/94 code, when found in persons over age 30, suggests that this pattern is highly resistant to change. In adolescent males, it is associated with delinquency.

With a correspondingly low 0, this code is likely to reflect a person with highly developed social techniques, who will use these skills to manipulate others. Thus, he or she may be involved in elaborate, antisocial "con" games. If Scale 3 is correspondingly high, it decreases the chance of acting out. In these cases, the expression of hostility is likely to be similar to that of the 34/43 code in that it will be indirect and often passive-aggressive. When Scale 6 is elevated along with Scales 4 and 9, extreme caution should be taken because these individuals will be very dangerous and have poor judgment. Their acting out will often be violent and bizarre, and will appear justified to themselves because of strong feelings of resentment toward others.

The most likely diagnosis is an antisocial personality, although caution should be made, especially when categorizing adolescents as these scales are more commonly elevated for both normal and abnormal adolescents. If Scale 8 is also high, it may reflect either a manic state or schizophrenia.

Personality and Interpersonal Characteristics

These individuals will often produce an external facade of being confident and secure, but underneath they will be immature, dependent, and insecure. They are likely to be narcissistic and incapable of deep emotional closeness. They will have a difficult time delaying gratification and often exercise poor judgment. Others will perceive them as being extraverted, talkative, uninhibited, restless, and needing emotional stimulation and excitement. Initially, they might make a good impression, but their antisocial style will soon become apparent. In particular, they will rationalize their own shortcomings and blame their problems on others.

Treatment Implications

There are numerous difficulties encountered in therapy with individuals having 49/94 code types. They have difficulty focusing for any length of time and are constantly embarking on often irrelevant tangents. Despite this, they can be quite articulate. They have difficulty delaying gratification and usually do not learn from experience but are more concerned with self-gratification (often at the expense of others). They are frequently irritable and, if confronted by a therapist, will express their fairly extensive

hostility. In addition, their typical coping strategy is through conning other people. Manipulation may involve a combination of charm laced with occasional belligerence. When this behavior occurs, it is advisable to confront it as soon as possible. Thus, treatment is likely to be slow, frustrating, and often unproductive. These individuals rarely volunteer for therapy but, rather, are referred by the court system or at the insistence of someone else (e.g., employer, spouse). External monitoring and motivation are usually required to keep them in treatment. However, because their anxiety level is quite low, they will not be motivated to change. Group treatment has been reported to be relatively helpful, and behavioral modification can often help them develop better coping styles. Despite this, termination is usually premature and associated with the client's feeling bored with the sessions, acting out, or a combination of the two.

68/86

Symptoms and Behaviors

The key features of people with the 68/86 code type are suspiciousness and distrustfulness, and they often perceive the intentions of others as suspect and questionable. They will be extremely distant from others, with few or no friends. They can be described as inhibited, shy, resentful, anxious, and unable to accept or appropriately respond to the demands that are made of them. As a result, they are highly involved in their fantasy world, uncooperative, and apathetic; and they have poor judgment and experience difficulty concentrating. Their sense of reality is poor, and they often experience guilt, inferiority, and mental confusion; sometimes their affect will be flat. The content of their thoughts can be expected to be unusual if not bizarre, frequently containing delusions of grandeur and/or self-reference. While their affect might be blunt, they are still internally quite anxious. Surprisingly, their past work history is often adequate provided the elevations on 6 and 8 are not extremely high. However, an intensification of their symptoms brought on by stress will usually disrupt their ability to work. Persons with this code are more often single and younger than 26 years of age. If they are married, their spouses are frequently also emotionally disturbed.

The most frequent diagnosis is paranoid schizophrenia, especially if Scale 4 is also elevated and 8 is relatively higher than 7. These persons will experience depression, inappropriate affect, phobias, and paranoid delusions. If Scale 7 is 10 points or more lower than Scales 6 and 8, this pattern is called the "paranoid valley" and emphasizes the presence of paranoid ideation. A highly elevated *F* with Scales 6 and 8 above 80 does not necessarily indicate an invalid profile. A paranoid state is also a frequent diagnosis with the 68/86 code; less frequently, organic brain disorders or severe anxiety disorders may be diagnosed.

Personality and Interpersonal Characteristics

Persons with this code type will be insecure with low self-confidence and poor self-esteem. Others perceive them as being unfriendly, negativistic, moody, and irritable. Because their level of social discomfort is high, they will feel most relaxed when alone and will generally avoid deep emotional ties. Their defenses will be poorly developed and, when under stress, are likely to regress (check the LSE/Low Self-Esteem and SOD/Social Discomfort scales).

Treatment Implications

Because a significant level of psychopathology is present with this profile, clinicians must be aware of a number of different issues related to further assessment and case management. In particular, treatment on either an inpatient or an outpatient basis needs to be decided. One of the major factors in this decision is a further assessment of the extent to which clients are a danger to themselves or others. A further consideration is whether psychopharmacological intervention and maintenance will help control psychotic thinking. In addition, basic daily living skills will be an issue. Clients might require training in basic social skills, assertiveness, job interviewing, and knowledge of resources to resort to when their symptoms increase. Insight-oriented therapy is often contraindicated as self-reflection might result in further regression. Instead, a concrete, behaviorally oriented method of intervention is likely to be more successful. One difficulty might be that these clients have unusual or even bizarre belief systems with quite different sets of logic than the therapist (check the BIZ/Bizarre Mentation scale). This might pose particular problems for attempts at cognitively based interventions. Furthermore, their level of suspicion and projection of blame will present further challenges. Because of their high level of mistrust, poor social skills, and social discomfort, they are likely to have difficulty forming a relationship with a therapist. Often, sessions will seem slow, unproductive, and characterized by long periods of silence. Impulsivity and regression are also likely to provide further treatment challenges.

69/96

Symptoms and Behaviors

Persons with 69/96 profiles are likely to be excited, oversensitive, mistrustful, energetic, and irritable. They may have difficulty thinking and exercise poor judgment. They feel extremely vulnerable to real or imagined threats and experience anxiety much of the time. Their typical response to stress is to withdraw into fantasy. They may have clear or subtle signs of a thought disorder including delusions, difficulty concentrating, hallucinations, tangential associations, incoherent speech, and appear perplexed and disoriented. They are likely to be obsessional, ruminative, and overideational. Diagnosis is likely to be either schizophrenia (paranoid type) or a mood disorder.

Personality and Interpersonal Characteristics

Individuals with this profile can be described as mistrustful and suspicious. They also have high needs for affection, and their relationships will often be passive-dependent. There is likely to be a clear discrepancy between how they describe themselves and how others perceive them. Whereas they describe themselves as calm, easygoing, happy, and in good health, others are likely to describe them as hostile, angry, and overreactive to even minor stress. These reactions to stress can result in their either becoming overly excited or apathetic and withdrawn. Thus, they have difficulties modulating their expression of emotions.

Treatment Implications

This code type is characteristic of inpatient populations. Psychopharmacological interventions to help control disorganized thinking or regulate mood can often be extremely

effective. Because of their disorganized, regressive, and ruminative thought processes, insight-oriented therapy is usually not effective. In addition, their lack of trust and suspiciousness often makes it difficult to form a therapeutic relationship. If a trusting relationship can be developed, concrete, problem-focused approaches are most effective.

78/87

Symptoms and Behaviors

The 78/87 code often occurs among psychiatric patients and reflects a level of agitation sufficiently intense to disrupt their daily activities. Usually, this profile represents a reaction to a specific crisis. They may have been previously functioning at an adequate level until some event or series of events triggered a collapse in their defenses ("nervous breakdown"). Their style of relating to others is passive, and they have difficulty developing and sustaining mature heterosexual relationships. They are lacking in self-confidence, often experience insomnia, and may have hallucinations and delusions. Common feelings include guilt, inferiority, confusion, worry, and fear, and they may have difficulties related to sexual performance.

The extent of elevations on Scales 7 and 8, and the relative heights between them, have important implications both diagnostically and prognostically. If Scale 7 is higher than Scale 8, the person's psychological condition is more susceptible to improvement and tends to be more benign. This has a tendency to be true regardless of the elevation of 8, as long as 7 maintains its relatively higher position. The higher Scale 7 suggests that the person is still actively fighting his or her problem and has some of his or her defenses still working. It also suggests an anxiety disorder rather than psychosis. Thus, ingrained bizarre thought patterns and withdrawn behavior have not yet become established. A relatively higher Scale 8, on the other hand, reflects more fixed patterns and is, therefore, more difficult to treat. This is particularly true if Scale 8 is over 75. If Scales 7 and 8 are both greater than 75 (with Scale 8 relatively higher), this suggests an established schizophrenic pattern, especially if the "neurotic triad" is low (check the BIZ/Bizarre Mentation scale). Even if schizophrenia can be ruled out, the condition tends to be extremely resistant to change, as, for example, with a severe, alienated personality disorder. If Scale 2 is also elevated, this raises the possibility of either a dysthymic or obsessive-compulsive disorder.

Personality and Interpersonal Characteristics

Persons with 78/87 code types are likely to feel inferior, inadequate, indecisive, and insecure. Their relationships will often be passive-dependent, and they will have difficulties asserting themselves in heterosexual relationships. They might be preoccupied in excessive and unusual sexual fantasies. They will feel extremely uncomfortable in most social relationships and are likely to defend themselves with excessive withdrawal (check SOD/Social Discomfort scale).

Treatment Implications

There may be a significant suicidal risk, which can be further evaluated by looking at the relative elevation of Scale 2, checking relevant critical items, taking a careful history, and asking relevant questions related to the client's thought processes.

89/98

Symptoms and Behaviors

The 89/98 code suggests persons who are highly energetic, perhaps to the point of hyperactivity. They will be emotionally labile, tense, and disorganized, with the possibility of delusions of grandeur sometimes with a religious flavor, especially if Scale 6 is also elevated. Their thought processes are likely to be tangential and speech bizarre, possibly characterized by neologisms, clang associations, and echolalia (check the BIZ/Bizarre Mentation scale). Their goals and expectations will be unrealistic; they often make extensive plans that are far beyond their ability to accomplish. Thus, their aspirations will be significantly higher than their actual achievements. Usually, they will have severe symptoms related to insomnia. Serious psychopathology is likely to be present.

The most frequent diagnosis is schizophrenia, or possibly a schizoaffective disorder with manic states. In addition, a severe personality disorder is a diagnostic consideration. Sometimes, the relative elevation of *F* can be used as an index of the relative severity of the disorder.

Personality and Interpersonal Characteristics

Their interpersonal relationships are childish and immature, and they will usually be fearful, distrustful, irritable, and distractible. Although they might be highly talkative and energetic, they will also prefer to withdraw from interpersonal relationships. They will resist any deep involvement with other people. While on the one hand they are grandiose and boastful, underneath they will have feelings of inferiority and inadequacy. When they do become involved with people, they demand considerable attention and become hostile and resentful when their needs are not met (check the ANG/Anger scale).

Treatment Implications

Because they are highly distractible and tangential, psychotherapeutic approaches with them are extremely difficult. Furthermore, their level of insight is poor, they resist psychological interpretations of their behavior, and they cannot focus on any one area for any length of time. A frequent defense is denial of any psychological problems along with grandiose thoughts and an inflated sense of their self-worth. Challenging these defenses is likely to provoke irritability, anger, or even aggression. If extensive delusions and hallucinations are present, antipsychotic medication may be indicated. Lithium may be useful if the mood component of their disorder predominates.

MMPI-2 CONTENT SCALES

One of the earliest efforts to develop a series of MMPI content scales was by Wiggins (1966, 1971) who organized scales based on an overall analysis of the contents of the MMPI items. He began with item clusters that were based on areas such as authority conflicts and social maladjustment. These clusters were revised and refined using factor analysis and evaluations of internal consistency. During the 1989 restandardization of the MMPI, many of the items relating to the Wiggins scales were altered or deleted. As a result, Butcher et al. (1990) developed a new set of 15 different content scales. At first,

provisional content scales were developed by rationally sorting the items into different content categories. These categories were then refined statistically by making item-scale correlations with psychiatric inpatients and correlations between the scales. Further validity studies have confirmed that they are at least as valid as the MMPI/MMPI-2/MMPI-A empirically derived scales (Barthlow et al., 1999; Ben-Porath, Butcher, & Graham, 1991; Ben-Porath et al., 1993; Butcher & Williams, 1992). A further advantage over the clinical scales is that they measure single dimensions. The practical significance is that they can be relatively easily interpreted using rational, intuitive strategies. In contrast, the MMPI clinical and validity scales are multidimensional. Thus, they require clinicians to work with them to extract the most useful and valid interpretations, often from a wide variety of possible descriptors.

An important function of the content scales is the ability to use them to refine the meanings of the clinical scales. For example, if an individual obtains an elevation on 4 (Psychopathic Deviance), clinicians can note possible corresponding elevations on FAM (Family Problems) and ASP (Antisocial Practices). If FAM is elevated but not ASP, the elevated 4 has more to do with family alienation and conflict than criminal and other forms of antisocial behavior. Thus, the content scales can incrementally increase the validity of the clinical scales (Barthlow et al., 1999; Ben-Porath et al., 1993).

In addition to clarifying the meanings of the scales, their interpretations and implications can also be extended. For example, elevations on 1, 2, and 3 are consistent with pain patients. However, in considering their prognosis for rehabilitation programs, it would also be important to assess their attitudes toward returning to work by noting the scores on WRK (Work Interference) and responsiveness to treatment by noting scores on TRT (Negative Treatment Indicators; M. Clark, 1996; Deardorff, 2000). Elevations above 65 on the content scales indicate that many of the descriptors for the scale apply to the person. Scales that are mildly elevated (60–64 inclusive) suggest that several of the behaviors apply to the person. Thus, the inclusion of the new MMPI-2 and MMPI-A content scales represents potentially important and easily interpreted dimensions of assessment. The content scales can be divided into the following clusters relating to internal symptoms, external aggression, negative self-views, and general problem areas.

Internal Symptomatic Behaviors

ANX/Anxiety Generalized anxiety, somatic difficulties, worries, insomnia, ambivalence, tension, a feeling that life is a strain, fear of losing his or her mind, pounding heart and shortness of breath, concentration problems, difficulties making decisions; symptoms clearly perceived and admitted to by the client.

FRS/Fears Multiple specific fears (nuisance animals, blood, dirt, leaving home, natural disasters, mice, snakes, etc.).

OBS/Obsessiveness Ruminates, difficulty with decision making, resistant to change, needless repetitive counting, may have compulsive behaviors such as counting or alphabetizing his or her experience; worried, sometimes overwhelmed by his or her own thoughts; others become easily impatient with the person. Persons with low scores are likely to be relaxed, secure, and unlikely to be depressed.

DEP/Depression High number of depressive thoughts, uninterested in life; feeling of emptiness; feeling of having committed unpardonable sins; cries easily; unhappy; possible suicidal ideation; sense that other people are not sufficiently supportive; sensitive to rejection, tense, passive feeling of hopelessness; helplessness about the future.

HEA/Health Concerns Numerous physical complaints regarding gastrointestinal, neurological, sensory, skin, cardiovascular and/or respiratory difficulties; problems of adjustment; worried and nervous; lacking in energy.

BIZ/Bizarre Mentation Psychotic thought processes, hallucinations (auditory, visual, olfactory), paranoid beliefs, strange thoughts, delusions.

External Aggressive Tendencies

ANG/Anger Difficulties in controlling anger, irritable, impatient, annoyed, stubborn, may swear; episodes of loss of control, possibly breaking objects or actually being physically abusive. Persons scoring low are unlikely to be depressed or have significant family problems.

CYN/Cynicism Distrust of other people; fear of being used, or that others will lie and cheat them; belief that the only reason for others not lying or cheating is fear of being caught; negativity toward friends and associates, belief that people are friendly only for selfish reasons. Persons with low scores might be highly achievement oriented.

ASP/Antisocial Practices Past legal and/or academic problem behaviors; expectation that others will lie, support of illegal behavior; enjoyment of criminal behavior of others; thought patterns that characterize criminal behavior, whether such behavior actually occurs or not. ASP has been found to be a better predictor (greater sensitivity and specificity) of antisocial personality disorder than *Pd* (Psychopathic deviance; S. Smith, Hilsenroth, Castlebury, & Durham, 1999) with a recommended cutoff of 55 or 60 (rather than the suggested cutoff of 65 implied by the MMPI-2).

TPA/Type A Driven, hardworking, competitive, hostile, irritable with time constraints, overbearing, annoyed with interruptions, tries to do more and more in less and less time, blunt and direct, petty regarding minor details (this scale is a better construct for use with males than females).

Negative Self-View

LSE/Low Self-Esteem Low self-confidence, feeling of insignificance, negative beliefs regarding self (clumsy, inept, unattractive), acutely aware of faults, feeling of being disliked by others, sometimes overwhelmed by his or her own faults, difficulty accepting compliments from others. Conversely, low scores suggest the person is secure, relaxed, and unlikely to be depressed.

General Problem Areas Cluster

SOD/Social Discomfort Shy, withdrawn, uneasy with others, introverted, dislikes social events, prefers to be alone. Persons with low scores are likely to be secure, relaxed, achievement oriented, assertive, and unlikely to be depressed or experience somatic symptoms.

FAM/Family Problems Family discord, unhappy childhood, difficult and unhappy marriages, families that do not express much love but are rather quarrelsome and unpleasant, possibly an abusive childhood.

WRK/Work Interference Personal difficulties that interfere with work; tension, worry, obsessiveness, difficulty concentrating, career indecision and/or dissatisfaction, poor concentration, dislike of coworkers; difficulty initiating work-related activities; little family support for career choice; easily defeated by difficulties.

TRT/Negative Treatment Indicators Dislike or distrust of helping professionals, discomfort in discussing difficulties, low level of self-disclosure, resistance to change, disbelief in the possibility of change, belief that no one can really understand or help them, preference for giving up rather than facing a crisis (M. Clark, 1996).

MMPI-A CONTENT SCALES

The MMPI-A content scales were developed and refined in much the same way as the MMPI-2 content scales. Some of the items were changed to be more relevant for adolescent populations. In addition, some new scales, such as the Adolescent-School Problems scale (instead of the adult WRK/Work Interference scale), were added, and others, such as the TPA (Type A) scale, were dropped because they were not considered relevant for adolescents. Elevations above 65 indicate that there has been extensive endorsement of the problems indicated in the scales whereas a mild elevation (60–64 inclusive) suggests that several of the descriptors apply to the person.

A-anx/Adolescent-Anxiety High scores suggest tension, nervousness, worry, sleep-related difficulties (nightmares, difficulty with sleep onset, early morning awakening); life feels like a strain; problems seem as if they are insurmountable; there will be feelings of impending doom, fears of losing his or her mind, confusion and difficulty concentrating, increase in family discord; girls in clinical settings report feeling depressed and have somatic complaints.

A-obs/Adolescent-Obsessiveness High scores suggest excessive worry, ruminations, obsessive counting of objects, extreme fear regarding making changes, difficulty making decisions, obsessing over past events or behaviors; others lose patience with them; boys in clinical settings are described as anxious, overly concerned with the future, dependent, worried, preoccupied, resentful, feel as if they deserve punishment; girls in clinical settings may have suicidal ideation and/or have actually made suicidal gestures.

A-dep/Adolescent-Depression High scores suggest fatigue, crying spells, self-criticism, feelings of being condemned and unworthy, feelings of hopelessness; life is uninteresting, suicidal ideation is present; there is difficulty initiating activities, dissatisfaction; boys in clinical settings might be further assessed for a history of abuse; girls in clinical settings have depression and low self-esteem; girls in school settings are likely to have poor grades, are unlikely to have noteworthy personal achievements, and are likely to be concerned about being overweight.

A-hea/Adolescent-Health Elevations indicate the presence of health problems that result in school absence and limit their physical activities; complaints cover several different physical areas including gastrointestinal (nausea, vomiting, constipation, stomach trouble), sensory problems (poor eyesight, hearing difficulty), neurological complaints (convulsions, paralysis, numbness, dizzy spells, fainting), cardiovascular problems (heart or chest pains), skin disorders, respiratory problems, excessive worry over health and belief that all related problems would be fine if their health difficulties could be solved; in clinical settings, they are likely to report being afraid of school; in school settings, they are likely to have academic and behavioral difficulties (school suspensions, course failures, low grades); girls in clinical settings are likely to report an increase in disagreements with parents; boys in clinical settings are described as anxious, worried, guilt prone, accident prone, perfectionistic (but less bright), clinging, fearful, and more likely to have lost weight.

A-aln/Adolescent-Alienation High scores indicate a high level of emotional distance, a feeling that no one really understands or cares for them, a sense that they are getting a raw deal from life, difficulty getting along with others, not liked, others are unkind and even out to get them; there is a belief that others have more fun than they do, low self-disclosure is likely; others interfere with their attempts to succeed; they feel anxious when talking to a group and are likely to have poor grades in school; girls may have a problem with weight gain; girls in clinical settings have few or no friends, increase in disagreements with parents; boys in clinical populations have low self-esteem and poor social skills.

A-biz/Adolescent-Bizarre Mentation High scores indicate very strange thoughts and experiences, possibly auditory, olfactory, and visual hallucinations, paranoid thoughts (plotted against, someone is trying to kill them), possible beliefs that evil spirits or ghosts are trying to control them; girls in clinical settings probably come from dysfunctional families, parents and/or siblings might have arrest records; boys in clinical settings are likely to have been under the supervision of a child protective worker, likely to exhibit bizarre and possibly psychotic behavior; individuals from school settings are likely to have numerous difficulties including poor grades, suspensions, and course failures.

A-ang/Adolescent-Anger High scores indicate that the person finds it difficult controlling anger, feels like breaking or smashing things, sometimes yelling to make a point and throwing tantrums to get his or her way; feels like getting into fist fights; shows irritability when others try to hurry him or her, impatient, especially likely to get into fights when drinking, likely to act out in school and/or home; adolescents in clinical

settings are extremely interested in violence and aggression, histories of assault, described as angry, resentful, impulsive, moody, externalize behaviors; boys in clinical settings are described as attention seeking, resentful, anxious, self-condemning but also dependent and clinging, may have a history of sexual abuse; girls in clinical settings are likely to be aggressive, delinquent, have been arrested, act out sexually (promiscuity), are flirtatious, wear provocative clothes, need to be supervised around boys.

A-con/Adolescent-Conduct Problems Elevations suggest that the client is oppositional, has legal problems, peer group is often in trouble; behavior problems including lying, stealing, shoplifting, swearing, vandalism; likely to enjoy other people's criminal behavior, might also enjoy making other people afraid of them; uses drugs and alcohol, has record of poor academic performance and school-related behavior problems (course failures, suspensions, lying and cheating), disobedient, impulsive; clinical girls are described as impulsive, angry, unpredictable, sexually active, provocative, resentful, impatient, require supervision around boys, unlikely to be depressed.

A-cyn/Adolescent Cynicism Persons scoring high are endorsing statements that they distrust other people. If other people are nice, it is only because they are trying to take unfair advantage of the people they are being nice to. Accordingly, high scorers feel guarded and misunderstood. Because others are out to get them and mainly concerned with self-interest, persons scoring high feel justified in having misanthropic attitudes. They may also believe that others are jealous of them.

A-lse/Adolescent-Low Self-Esteem High scores indicate that the individual feels unattractive, useless, has little ability, many faults, low self-confidence, unable to do anything particularly well including planning own future, confused and forgetful, difficulty accepting compliments, susceptible to social pressure, passive; high-scoring boys should be further assessed for the possibility of sexual abuse; girls are likely to report weight gain, poor grades, and no noteworthy personal achievements; boys in clinical settings are described as having poor social skills; girls in clinical settings will be depressed, are likely to have learning disabilities, have increasing numbers of conflicts with their parents, suicidal thoughts, and possibly suicidal gestures.

A-las/Adolescent-Low Aspirations High scores indicate a low level of interest, especially academically; the person dislikes studying, reading, listening to lectures (especially science); has problems initiating activities, gives up easily, dislikes facing difficult situations; has low expectations for achievement and little interest in continuing on to college; described by others as lazy, has poor grades, little interest in school activities; clinical girls are likely to report sexual acting out, very unlikely to report having won a prize or award; clinical boys are likely to have been truant in school and run away from home.

A-sod/Adolescent-Social Discomfort High scores indicate that the person is shy, prefers to be alone, difficulty making friends, extremely uncomfortable when addressing a group, dislikes parties and crowds, difficult to get to know, uncomfortable meeting new people, dislikes initiating conversations, might actively avoid others, unlikely to report

using drugs or alcohol; boys are likely to avoid school activities; girls in clinical settings are unlikely to be involved in acting out, are uninterested in boys, have few friends, may be depressed, have eating difficulties, may be fearful, withdrawn, physically weak, and are not likely to be involved with drugs, alcohol, or irresponsible behavior.

A-fam/Adolescent-Family Problems High scorers are likely to have extensive difficulties with parents and other family members including fault-finding, jealousy, little love, serious arguments, poor communication; they long for the day when they can finally leave home, feel that parents punish them unfairly, show little acceptance of responsibility around home, feel that they cannot depend on their family in times of need; beatings and runaways are possible, however, problems usually do not extend into the legal justice system; there may be some school-related difficulties (low grades, suspensions); may reflect marital difficulties of parents; girls in school settings report possible exam failure and/or weight gain; in clinical settings there may be more externalizing behaviors including lying, cheating, stealing as well as somatic complaints, crying, guilt, timidity, and withdrawal; boys in clinical settings are described as sad, secretive, uncommunicative, disliked, self-conscious, unloved, dependent, resentful, attention seeking, and self-blaming; girls in clinical settings are typically described as immature, likely to fight, cruel, destructive, secretive, self-conscious, hyperactive, provocative, sexually acting out (promiscuity), and preoccupied with sex; further assessment should include possible sexual abuse for girls and possible physical abuse for boys.

A-sch/Adolescent-School High scores indicate a wide number of school-related difficulties including low grades, truancy, easily upset by school events, learning disabilities, low level of social competence, boredom, suspensions, dislike of school, disciplinary actions, difficulty concentrating, probations, and negative attitudes toward teachers; feels that school is a waste of time; often school-related difficulties are specific to school itself and do not spill over into other areas; boys from clinical populations are likely to have run away, been irresponsible, and have a history of drug use, particularly amphetamine, they should be further evaluated for the possibility of sexual abuse; girls from clinical populations may have learning disabilities and/or academic underachievement.

A-trt/Adolescent-Negative Treatment Indicators High scores indicate negative attitudes and feelings toward health care professionals; they do not like to share personal information with others; they feel that they can never really be understood and others do not really care what happens to them; they will have anxiety related to people asking them personal questions; they have difficulty planning for the future and are unwilling to take responsibility for the negative things in their lives; they feel that they have many secrets they need to keep to themselves.

HARRIS-LINGOES AND *SI* SUBSCALES

One of the more popular developments has been the reorganization by Harris and Lingoes (1955/1968) of the standard scales into more homogeneous content categories.

These subscales were constructed by intuitively grouping together items that seemed to reflect single traits or attitudes contained in the already existing MMPI Scales 2, 3, 4, 6, 8, and 9. Ben-Porath et al. (1989) further developed subscales similar to the Harris-Lingoes subscales for Scale 0. No subscales were developed for 1 and 7 because these were considered to be relatively homogeneous in their item content. These same subscales have been carried over for use with the MMPI-A. The subscales and a brief summary of the meanings associated with high scores are provided in this section. These summaries are derived from material by Harris and Lingoes (1968), and extensions of these materials as summarized by Butcher et al. (1990), Butcher and Williams (1992), J. Graham (2000), Greene (2000), and Levitt and Gotts (1995). Scoring templates and profile sheets for the MMPI-2 and MMPI-A Harris-Lingoes subscales are available from National Computer Systems.

Although the Harris and Lingoes subscales show high intercorrelations with the parent scales (Harris & Lingoes, 1968) and relevant code types (McGrath, Powis, & Pogge, 1998), the internal consistency of the subscales is somewhat low (.04 to .85; Gocka, 1965). Several initial validity studies are available (Boerger, 1975; Calvin, 1975; N. Gordon & Swart, 1973) that demonstrate the potential clinical usefulness of these subscales. The Social Introversion subscales have been found to account for 90% of the variance of the *Si* scale, and convergent and discriminant validity was demonstrated based on an analysis of spouses' ratings of each other (Ben-Porath et al., 1989). The practical importance of both sets of subscales is that they provide a useful supplement for interpreting the original scales. For example, a clinician can assess whether a person scoring high on Scale 4 (Psychopathic Deviate) achieved that elevation primarily because of family discord (*Pd* 1), authority problems (*Pd* 2), or social imperturbability (*Pd* 3). This breakdown is likely to be quite helpful in interpreting why a client received a high score that was unexpected based on the person's history. It might also be quite useful in interpreting the significance associated with moderate elevations ($T = 60–65$). A further situation to score and interpret the Harris-Lingoes scales is to understand the possible reasons for contradictory descriptions such as might emerge if both Scales 2 and 9 were elevated. However, if the clinical scales are either in the normal range, or quite high, the Harris-Lingoes scales are not particularly useful. Only Harris-Lingoes and *Si* subscale elevations of $T = 65$ or greater should be interpreted.

The Harris-Lingoes and *Si* subscales should not be used for routine interpretations because they are quite time consuming to hand-score. Rather than scoring all the Harris-Lingoes and *Si* subscales, clinicians can select and score only those that are relevant for refining and clarifying the meanings of clinical scales that are in question. Despite some validity efforts, the amount of research available is still inadequate, and, in many cases, the internal consistency of the subscales is insufficient. Thus, any interpretations should be made cautiously and be considered as hypotheses in need of further support. This is particularly true for the MMPI-A, in which there has been even less investigation using the Harris-Lingoes and *Si* subscales than for the MMPI/MMPI-2. Furthermore, item deletions and alterations between the MMPI/MMPI-2 and MMPI-A, primarily for the *Si* scale, bring into question the transferability of the Harris-Lingoes and *Si* scales with the adolescent version of the MMPI.

Scale 2. Depression

D1/Subjective Depression Unhappy, low energy, sense of inferiority, low self-confidence, socially uneasy, few interests.

D2/Psychomotor Retardation Low energy, immobilized, socially withdrawn, listless.

D3/Physical Malfunctioning Reports wide variety of physical symptoms, preoccupied with health, denial of good health.

D4/Mental Dullness Low energy, pessimistic, little enjoyment of life; difficulties with concentration, attention, and memory; apathetic.

D5/Brooding May feel as if he or she is losing control of his or her thoughts; broods, cries, ruminates, feels inferior, and is hypersensitive.

Scale 3. Hysteria

Hy1/Denial of Social Anxiety Extraverted, comfortable with social interaction, minimally influenced by social standards.

Hy2/Need for Affection Strong needs for affection with fears that these needs will not be met, denies negative feelings toward others.

Hy3/Lassitude-Malaise Subjective, discomfort, poor health, fatigued, poor concentration, insomnia, unhappiness.

Hy4/Somatic Complaints Wide variety of physical complaints, denial of hostility toward others.

Hy5/Inhibition of Aggression Denial of hostility and anger, interpersonally hypersensitive.

Scale 4. Psychopathic Deviate

Pd1/Familial Discord Family that was critical, unsupportive, and interfered with independence.

Pd2/Authority Conflict Rebellion against societal rules, beliefs of right/wrong that disregard societal norms, legal/academic difficulties.

Pd3/Social Imperturbability Opinionated, socially confident, outspoken.

Pd4/Social Alienation Isolated from others, feels poorly understood.

Pd5/Self-Alienation Unhappy with self, guilt and regret regarding past behavior.

Scale 6. Paranoia

Pa1/Persecutory Ideas Perceives world as dangerous, feels poorly understood, distrustful.

Pa2/Poignancy Feels lonely, tense, hypersensitive, possibly high sensation-seeking.

Pa3/Naivete Overly optimistic, extremely high moral standards, denial of hostility.

Scale 8. Schizophrenia

Sc1/Social Alienation Feels unloved, mistreated, and possibly persecuted.

Sc2/Emotional Alienation Depression, fear, possible suicidal wishes.

Sc3/Lack of Ego Mastery, Cognitive Strange thoughts, sense of unreality, poor concentration and memory, loss of mental control.

Sc4/Lack of Ego Mastery, Conative Depressed, worried, fantasy withdrawal, life is too difficult, possible suicidal wishes.

Sc5/Lack of Ego Mastery, Defective Inhibition Sense of losing control of impulses and feelings, labile, hyperactive, cannot control or recall certain behaviors.

Sc6/Bizarre Sensory Experiences Hallucinations, peculiar sensory and motor experiences, strange thoughts, delusions.

Scale 9. Hypomania

Ma1/Amorality Selfish, poor conscience, manipulative; justifies amoral behavior by believing others are selfish and opportunistic.

Ma2/Psychomotor Acceleration Restless, hyperactive, accelerated thoughts and behaviors, seeks excitement to reduce boredom.

Ma3/Imperturbability Unaffected by concerns and opinions of others, denies feeling socially anxious.

Ma4/Ego Inflation Unrealistic perception of abilities, resentful of demands placed on himself or herself.

Scale 0. Social Introversion

(*Note:* The Social Introversion subscales are scored on the MMPI-2 Supplementary Scales Profile sheet; scoring templates are available from National Computer Systems.)

Si1/Shyness Easily embarrassed, reluctant to initiate relationships, socially uncomfortable, shy.

Si2/Social Avoidance Dislike and avoidance of group activities, parties, social activities.

Si3/Self/Other Alienation Poor self-esteem, self-critical, low self-confidence, sense of ineffectiveness.

CRITICAL ITEMS

An alternative to content analysis, other then scoring and interpreting actual scales, is to interpret the meanings of single items or clusters of items that seem, based on their content, to relate to different areas of psychopathology (depressed suicidal ideation, mental confusion, etc.) or direction on these items could represent serious pathology, regardless of how the person responded on the remainder of the inventory. These items have been referred to as *pathognomonic items, stop items,* or, more frequently, *critical items.* It has been assumed that the direction in which a person responds represents a sample of the person's behavior and acts like a short scale that indicates the client's general level of functioning. A listing of critical items can be found in the MMPI-2 manual, and they are typically scored by most computer-assisted programs. The critical items will be most useful if clinicians look at the individual item content in relationship to the specific types of information that the item reveals. This information might be used to guide further interviewing. In addition, the items themselves, along with the responses ("True" or "False") might be included in the psychological report to provide qualitative information regarding the client. However, some caution should be taken in their interpretation, as they are both subject to an acquiescing response set (most items are keyed in the "True" direction) and faking bad. They should not be considered to be scales but rather direct communications to the clinician about areas specific to the item content.

While lists of critical items have been included in standard interpretive guides for the MMPI-A (Archer, 1992a; Butcher & Williams, 1992), clinicians should use these lists with adolescents with caution. First, normal adolescents as well as clinical populations of adolescents endorse, on average, twice the number of critical items as normal adults (Archer & Jacobson, 1993). In addition, normal adolescents and clinical populations endorse item frequencies about equally, thereby suggesting that the items themselves should not be used to differentiate between these two groups. This means that empirical attempts to develop critical item lists for adolescents might be quite difficult. As for the MMPI/MMPI-2, clinicians should not treat the different clusters of critical items as rough scales to be interpreted. Rather, the individual item content should be used to develop specific interview questions, and the relative deviancy of these items should be handled with appropriate tolerance.

MMPI-2 AND MMPI-A SUPPLEMENTARY SCALES

Since the initial publication of the MMPI, more than 450 new scales have been developed (see Levitt & Gotts, 1995). Some of these have been developed for normals and are unrelated to pathology, such as dominance (*Do*) and social status (*St*). Other scales relate

more directly to pathological dimensions, and often use the data from Hathaway and McKinley's original standardization sample or the more recent restandardization group. Scoring is possible only if the entire 567 MMPI-2 or 478 MMPI-A items are given. Although exact cutoffs for determining high scores have not been specified, they are generally $T = 65$. Scoring templates and profile sheets are available through National Computer Systems. The scales selected for inclusion on this profile sheet are considered most useful, have been most extensively researched, or show promise in terms of future usefulness and/or are likely to be researched more extensively in the future. The following lists provide the names and interpretations surrounding scale elevations.

MMPI-2 Supplementary Scales

A/Anxiety High scores indicate that the person is upset, shy, retiring, insecure, has low self-confidence, is inhibited, uncertain, hesitant, conforming, under stress, and has extreme difficulty making decisions. Low scores indicate that the individual is extraverted, secure, relaxed, energetic, competitive, and generally has an absence of emotional difficulties.

R/Repression High scorers tend to be submissive, overcontrolled, slow, clear thinking, conventional, formal, cautious, use denial and rationalization, and go to great lengths to avoid unpleasant interpersonal situations. Low scorers are likely to be dominant, enthusiastic, excitable, impulsive, self-indulgent, outspoken, and achievement oriented.

Es/Ego Strength This scale assesses the degree to which a client is likely to benefit from psychotherapy, but it is probably specific to predicting the response of neurotic patients to insight-oriented therapy; it is probably not useful for other types of patients or other kinds of treatments (J. Graham, 1978). High scores suggest these persons can benefit from psychotherapy because they are likely to be adaptable and possess personal resources, have good reality contact, are tolerant, balanced, alert, have a secure sense of reality, will seek help in situational difficulties, possess strongly developed interests, are persistent, can deal effectively with others, have a sense of personal adequacy, can easily gain social acceptance, and have good physical health. Low scores reflect general maladjustment. These people are likely to have low self-esteem, a poor self-concept, lack personal resources, feel insecure, be rigid and moralistic, have chronic physical problems, possess fears and phobias, are confused and helpless, have chronic fatigue, may be withdrawn and seclusive, inhibited, have personality rather than situational problems and poor work histories, and will, therefore, have difficulty benefiting from psychotherapy.

MAC-R/MacAndrew Alcoholism Scale-Revised The MAC-R scale is best considered a measure of the potential for substance abuse. It differentiates between outpatient alcoholics and nonalcoholic psychiatric outpatients, identifies persons who are at risk of later developing alcohol-related problems. The *potential* to become involved in alcohol use is assessed rather than current alcohol use. In addition, the scale has difficulty differentiating alcohol abusers from other substance abusers. High scores on the MAC-R scale primarily suggest actual or potential substance abuse but may also suggest extraversion,

affiliation, confidence, assertiveness, risk taking, sensation seeking, past school behavior problems, the possibility of having experienced blackouts, and possible difficulties with concentration. Low scores are not only a contraindication of substance abuse, but also may suggest introversion, conformity, and low self-confidence. If low scores in a known substance abuser do occur, this suggests that the abuse is based more on psychological disturbance than typical addictive processes. The recommended *raw* score cutoff to indicate the initial point of drug and/or alcohol problems for males is 26 to 28, whereas for females it is a lower 23 to 25. The MAC-R is not particularly effective with African Americans and other non-Caucasian respondents. High scorers are likely to be extraverted, impulsive risk takers who will benefit from a group-oriented, confrontive treatment approach. Low scorers are more likely to be introverted, withdrawn, depressed risk avoiders who will be more likely to benefit from a supportive and relatively nonconfrontational treatment approach.

AAS/Addiction Acknowledgment Scale High scores suggest a conscious awareness of and willingness to share information related to drug and/or alcohol-related problems. It is the most sensitive MMPI-2 scale for detecting substance abuse (Rouse, Butcher, & Miller, 1999; L. Stein, Graham, Ben-Porath, & McNulty, 1999). Low scores merely clarify that the person has not acknowledged these problems (although there is still the possibility that they do have drug and/or alcohol-related difficulties).

APS/Addiction Potential Scale High scores indicate that the person has a considerable number of lifestyle and personality factors consistent with those who abuse alcohol and/or drugs. The scale does not necessarily measure the extent of current use but more the potential for developing such problems. This means that if the APS (or MAC-R) is used to identify persons who are actually abusing substances, it is likely to result in a high number of false positives (Rouse et al., 1999). If the person scores in the normal-to-low range but history reveals that they have a drug and/or alcohol problem, this problem is probably based primarily on psychological maladjustment (drug/alcohol use as self-medication) rather than a typical addictive pattern (harmful habits, peer group issues, physiological impact of the drug). This scale is quite similar to the MAC-R scale, but it has used more of the newer MMPI-2 item pool than the MAC-R. There is some indication that it measures the same factors as the MAC-R and may do so either as effectively (Rouse et al., 1999; L. Stein et al., 1999) or more effectively (Greene, Weed, Butcher, Arredono, & Davis, 1992; Weed, Butcher, Ben-Porath, & McKenny, 1992).

MDS/Marital Distress Scale High scores indicate the person is experiencing marital distress; this scale is more specifically related to marital difficulties than either the FAM content scale or Scale 4 (both of which assess relationship difficulties not necessarily specific to marriage); MDS should be interpreted only for persons who are married, separated, or divorced.

O-H/Overcontrolled Hostility Scale High scores suggest that the person is emotionally constricted, bottles up anger, and may overreact, possibly becoming physically or verbally aggressive; the aggressiveness usually occurs as rare incidents in a person

who is otherwise extremely well controlled; the scale is most useful in understanding past behavior rather than predicting the likelihood of future hostility; some persons who score high are not actively struggling to control dangerous hostility but are very well controlled and highly socialized. Thus, the scale is more directly a measure of persons who deny aggressive actions and are somewhat constricted; therapy, at least initially, might seem superficial and lacking in affect.

Do/Dominance Elevations indicate that the individual is self-confident, realistic, task oriented, feels a sense of duty toward others; is competent to solve problems, socially dominant, poised, and self-assured in working with groups; takes the initiative in relationships, possesses strong opinions, perseveres at tasks, and has a good ability to concentrate; the scale is useful and frequently used in personnel selection (e.g., police officer selection).

Re/Responsibility High scores suggest that the individual possesses high standards, a strong sense of justice and fairness, strong (even rigid) adherence to values, is self-confident, dependable, trustworthy; the scale is a general index of positive personality characteristics; often useful in personnel screening.

Mt/College Maladjustment High scores indicate general maladjustment among college students; they are likely to be worried, anxious, and procrastinate; they are pessimistic, ineffectual, somatize stress, and feel that, much of the time, life is a strain.

GM/Masculine Gender Role Persons who score high (both males and females) are likely to be self-confident, deny feeling afraid or worried, and be persistent in pursuing their goals; females scoring high are likely to be honest, unworried, and have a willingness to explore new things; high scores on GM with correspondingly low scores on Gf indicate stereotypic male interests and orientations; high scores on both GM and Gf suggest androgyny (the person has both masculine and feminine characteristics); low scores on GM along with high scores on Gf suggest stereotypic feminine interests and orientation; low scores on both scales suggest an undifferentiated masculine/feminine orientation; this is still an experimental scale in need of further research.

GF/Feminine Gender Role High scores suggest the endorsement of stereotypically feminine interests and orientations, and may also suggest religiosity and possibly abuse of alcohol and/or nonprescription drugs; males scoring high may be hypercritical, express religiosity, avoid swearing but act bossy, and have a difficult time controlling their temper. This is still an experimental scale in need of further research.

PK/Posttraumatic Stress Disorder Scale High scores indicate emotional distress, depression, anxiety, sleep disturbances, guilt, loss of control over thinking, a feeling of being misunderstood and mistreated by others; the scale does not determine that trauma has actually occurred but indicates that the symptoms reported are consistent with persons exposed to traumatic events; the existence of a trauma still needs to be determined through other means.

PS/Posttraumatic Stress Disorder Scale This is a second trauma-related scale that is listed on the supplementary scale profile sheet but is currently an experimental scale under development. (*Si1, Si2, Si3*/Shyness, Social Avoidance, and Self-Other Alienation are listed here for clarification; they occur on the supplementary scale profile sheet but actually refer to the Social Introversion content subscales described in the previous section on Harris-Lingoes and *Si* subscales.)

MMPI-A Supplementary Scales

MAC-R/MacAndrew Alcoholism Scale High scores suggest that the person is similar to others who have alcohol or drug problems; dominant, assertive, egocentric, self-indulgent, impulsive, unconventional; risk taker and sensation seeker; increased possibility of conduct disorder and legal difficulties. Low scores suggest that the person is dependent, conservative, avoids sensation-seeking activities, is overcontrolled, and indecisive.

ACK/Alcohol Drug Acknowledgment Scale Persons who score high have a conscious awareness of and willingness to admit to alcohol- and/or drug-related problems; includes problem use, reliance on alcohol to cope or as a means of freely expressing feelings, harmful substance abuse habits; friends or acquaintances may tell them that they have alcohol and/or drug problems; they may get into fights while drinking.

PRO/Alcohol Drug Proneness Scale A high score suggests that the person is prone to developing drug- and/or alcohol-related problems, school and home behavior problems. No obvious items related to drugs and alcohol are included on the scale; therefore, the scale measures personality and lifestyle patterns more consistent with alcohol- and drug-related problems. The scale does not so much measure current alcohol or drug use patterns (although they may still be present; quite similar to the MMPI-2 APS scale).

IMM/Immaturity Scale High scorers are untrustworthy, undependable, boisterous; quickly become angry, are easily frustrated, may tease or bully others; are resistant, defiant, and are likely to have a background of school and interpersonal difficulties.

A/Anxiety General maladjustment, anxiety, distress, emotionally upset, experiences discomfort.

R/Repression Submissive, conventional, works hard to avoid unpleasant or disagreeable situations.

RECOMMENDED READING

Archer, R. A. (1992). *MMPI-A: Assessing adolescent psychopathology.* Hillsdale, NJ: Erlbaum.

Butcher, J. N. (1999). *A beginner's guide to the MMPI-2.* Washington, DC: American Psychological Association.

Butcher, J. N., & Williams, C. L. (1992). *Essentials of MMPI-2 and MMPI-A interpretation.* Minneapolis, MN: University of Minnesota Press.

Caldwell, A. (2001). What do the MMPI scales fundamentally measure? Some hypotheses. *Journal of Personality Assessment, 76,* 1–17.

Finn, S. E. (1996). *A manual for using the MMPI-2 as a therapeutic intervention.* Minneapolis, MN: University of Minnesota Press.

Friedman, A., Lewak, R., Nichols, D., & Webb, J. F. (2000). *Psychological assessment with the MMPI-2.* Mahwah, NJ: Erlbaum.

Graham, J. R. (2000). *MMPI-2: Assessing personality and psychopathology* (3rd ed.). New York: Oxford University Press.

Lewak, R. W., Marks, P. A., & Nelson, G. E. (1990). *Therapist guide to the MMPI and MMPI-2.* Muncie, IN: Accelerated Development.

Chapter 8

MILLON CLINICAL MULTIAXIAL INVENTORY

The Millon Clinical Multiaxial Inventory (MCMI) is a standardized, self-report questionnaire that assesses a wide range of information related to a client's personality, emotional adjustment, and attitude toward taking tests. It has been designed for adults (18 years and older) who have a minimum of an eighth-grade reading level. The MCMI is one of the few self-report tests that focus on personality disorders along with symptoms that are frequently associated with these disorders. Originally developed in 1977 (Millon, 1977), it has since been through two revisions (MCMI-II; Millon, 1987; MCMI-III; Millon, 1994, 1997). Since its original publication, it has stimulated more than 600 published papers on or using it and has become one of the more frequently used tests in clinical practice (Camara et al., 2000; C. Piotrowski & Zalewski, 1993; Watkins et al., 1995). Indeed, it is one of the few tests that has "risen through the ranks" of test usage over the past 30 years. Among objective personality tests for clinical trainees to be familiar with, the MCMI was ranked by directors of clinical training programs second only to the MMPI/MMPI-2 in importance (C. Piotrowski & Zalewski, 1993). Its popularity is further supported by its use in several different countries and its translation into a number of different languages.

The current version, the MCMI-III, is composed of 175 items that are scored to produce 28 scales divided into the following categories: Modifying Indices, Clinical Personality Patterns, Severe Personality Pathology, Clinical Syndromes, and Severe Syndromes (see Table 8.1). The scales, along with the items that comprise the scales, are closely aligned to both Millon's theory of personality and the *DSM-IV* (1994). For example, an item endorsing a person's belief in his or her own superiority would be part of the Narcissistic scale, because the content clearly relates to components of Millon's and the *DSM-IV*'s conceptualization of the narcissistic personality. Many of the scales have both theoretical and item overlap—an important fact to keep in mind when conceptualizing the client and interpreting the scales. Thus, an elevation on both the Antisocial and Sadistic scales would reflect a person who has sadistic features along with legal difficulties and impulsiveness, and who is interpersonally exploitive. Similarly, a person scoring high on the Antisocial scale might have a corresponding elevation on the Alcohol Dependence scale. The corresponding elevations on conceptually related scales allow for a more complete understanding of the client.

In some ways, the MCMI is an alternative or even a competitor to the MMPI. Both instruments cover a wide range of adult pathology that assess both long-standing personality patterns as well as clinical symptomatology. In other ways, the MCMI nicely

Table 8.1 MCMI-III Scale categories, abbreviations, number of items, and reliabilities

Scale Category/Name	Abbreviation	No. of Items	Alpha
Modifying Indices			
Disclosure	X	NA	NA
Desirability	Y	21	.85
Debasement	Z	33	.95
Validity	V	4	NA
Clinical Personality Patterns			
Schizoid	1	16	.81
Avoidant	2A	16	.89
Depressive	2B	15	.89
Dependent	3	16	.85
Histrionic	4	17	.81
Narcissistic	5	24	.67
Antisocial	6A	17	.77
Aggressive (Sadistic)	6B	20	.79
Compulsive	7	17	.66
Passive-Aggressive (Negativistic)	8a	16	.83
Self-Defeating	8B	15	.87
Severe Personality Pathology			
Schizotypal	S	16	.85
Borderline	C	16	.85
Paranoid	P	17	.84
Clinical Syndromes			
Anxiety	A	14	.86
Somatoform	H	12	.86
Bipolar:Manic	N	13	.71
Dysthymia	D	14	.88
Alcohol Dependence	B	15	.82
Drug Dependence	T	14	.83
Posttraumatic Stress Disorder	PT	16	.89
Severe Syndromes			
Thought Disorder	SS	17	.87
Major Depression	CC	17	.90
Delusional Disorder	PP	13	.79

Source: Adapted from *Millon Clinical Multiaxial Inventory-III Manual* (2nd ed.), by Millon, 1997, Mineapolis: National Computer Systems.

complements the MMPI as the MMPI focuses primarily on Axis I disorders whereas the MCMI was specifically designed to assist in diagnosing Axis II disorders. One important advantage of the MCMI is that it is considerably shorter than the MMPI-2 (175 vs. 567 items) and yet provides a wide range of information. The MCMI takes only 20 to 30 minutes to complete; however, the research base, validity studies, and options for interpretations are clearly more extensive for the MMPI than for the MCMI. Neither instrument should be considered to provide diagnosis. Instead, they provide considerable

information relevant to diagnosis. In this sense, they place the clinician in the right "diagnostic ballpark," but the clinician must then integrate this with other information to make the final diagnosis. In other words, tests (or computer reports) don't diagnose (or make decisions); only practitioners can perform this function.

Factors that greatly assist in useful interpretation are familiarity with the theoretical constructs as well as experience with relevant clinical populations. Theoretical knowledge can be greatly assisted through familiarity with Millon and Davis's (1996) *Disorders of Personality* as well as the diagnostic criteria of the *DSM-IV.* This emphasis on clinical populations also focuses on the principle that the MCMI is intended for psychiatric populations and should not be used with normal persons or those who are merely mildly disturbed. Interpretations should be restricted to persons who scored at or above the designated cutoff scores (75 and 85). Practitioners should resist the temptation to attempt interpretations of persons who have mild "elevations" on the scale but who are still clearly below the formal cutoff.

HISTORY AND DEVELOPMENT

Development of the Original MCMI

Shortly after Millon published his 1969 text, *Modern Psychopathology,* fellow professionals urged him to develop an instrument that would operationalize and measure the dimensions of personality as outlined in the book. By 1972, an initial form was developed: the Millon–Illinois Self-Report Inventory (MI-SRI). Over the next five years, the items were further developed, refined, and coordinated with the upcoming personality disorders that were later to be incorporated into *DSM-III* (1980). When the initial refinements were completed, the test was published and renamed the Millon Clinical Multiaxial Inventory (MCMI; Millon, 1977).

The formal development of the original MCMI used a combination of rational theory-based, as well as empirical, procedures. The first step was the development of a large pool of face valid questions—a total of 3,500 items derived from Millon's (1969) theories. These were then rationally grouped into 20 different scales. The number of items was initially reduced by the test developers by rewording those that were poorly worded and removing those that were redundant. Further refinement was done empirically by having patients rate the clarity and difficulty of the items. A further procedure involved having clinicians regroup the items into scales to evaluate the extent to which these scales related to those originated by the test developers. Based on these procedures, the items were then grouped into two equivalent provisional research forms, with 556 items in each form. The forms were administered to 200 patients, and their responses were evaluated for their endorsement frequency and item-scale intercorrelations. The highest within-scale item intercorrelations were retained, and items that were either very frequently (> 85%) or very rarely (15%) endorsed were eliminated. The research form was thereby reduced to a test composed of 289 items.

The 289-item research form was given to 167 clinicians who blindly rated 682 of their patients on 20 different variables after having given them the form. The amount of endorsement frequency and the degree of scale overlap were then used to reduce the

items from 289 to 154. Based on this initial validation procedure, three scales were dropped (Sociopathy, Hypochondriasis, and Obsession-Compulsion), and three new scales were developed and added (Drug Abuse, Alcohol, and Hypomania). This brought the total number of surviving items to 175, with 733 different keyings on the 20 different scales.

The scales were initially standardized on 1,591 clinical subjects used in the construction phase of the test. This sample was used to establish the optimal cutoff scores for determining the presence or absence of certain characteristics. A group of 297 nonclinical subjects was used to establish the responses of a normal comparison group. In 1981, the MCMI responses of 43,218 patients were reviewed to further refine and recalculate the cutoff scores.

One feature of the MCMI and its revisions is the use of cutoffs related to Base Rate (BR) scores to designate the presence or absence of a particular characteristic. The BR score, like the more familiar *T* score, is essentially a means of transforming a raw score into a more meaningful score for interpretation. However, BR scores are derived from the percentage of a population that has been deemed to have a certain characteristic or syndrome. For example, 17% of a psychiatric population can be considered to have clear characteristics of a dependent personality whereas only 1% is considered to have clear features of a sadistic personality. This means that decisions regarding client characteristics are made when a client scores in a range that is consistent with either of these two syndromes. However, the relatively more frequent psychiatric disorders with high BRs (i.e., Dependent) require relatively lower cutoff points than those rare disorders with low BRs (i.e., Sadistic). Millon arbitrarily set a BR score of 85 to indicate that the characteristic(s) in question was definitely present. A lower BR score of 75 indicated that some of the features were present. Additional cutoff or anchor points were set at 35 to represent the median score for normal or nonpsychiatric groups, and at 60, the median for psychiatric populations. This BR approach has been theoretically encouraged by a number of authors (Finn, 1982; Widiger & Kelso, 1983) and empirically demonstrated to increase diagnostic accuracy when compared with the more frequently used *T* score approach (Duthie & Vincent, 1986).

Development of the MCMI-II

The MCMI-II (Millon, 1987) maintained most of the features of the original MCMI. Its development was motivated by a need to incorporate additional research and theory on personality disorders while remaining aligned with the criteria outlined in *DSM-III* and *DSM-III-R*. In addition, between 40 and 50 of the original MCMI items were found to be expendable. Items were developed for two new scales, in part by dividing the previous Negativistic scale into separate scales for Passive-Aggressive (Negativistic) and Self-Defeating. Similarly, the earlier Antisocial-Aggressive scale was divided into an Antisocial scale and an Aggressive/Sadistic scale. Additional items were generated with procedures similar to those used for the original MCMI. This resulted in an MCMI-II Provisional Form of 368 items, which was given to 184 patients who had been carefully diagnosed using *DSM-III-R* (1987) criteria. Items were retained or deleted based on the extent to which they could differentiate relevant diagnostic criterion groups. Like the earlier MCMI, the MCMI-II totaled 175 items, but they were keyed on 22 (as opposed to only 20) different scales. In an attempt to reduce scale

intercorrelation, individual items were given weightings of 1, 2, or 3 points, based on their relative importance for the specific scales they were being keyed on. Optimal BR cutoff scores were based on a standardization group of 1,292 patients who had a wide variety of presenting problems.

Development of the MCMI-III

Ongoing research, new conceptual developments, and the publication of the *DSM-IV* contributed to the MCMI-II's revision into its latest version, the MCMI-III (Millon, 1994, 1997). With procedures similar to those used for the MCMI and MCMI-II, a provisional 325-item test was developed; Depressive and PTSD scales were added. The Self-Defeating and Sadistic Personality Disorder scales were maintained, although these diagnoses were eliminated from the *DSM-IV*. The final MCMI-III still totaled 175 items, but 90 of the items from the MCMI-II were "changed" (85 remained the same). Actually, most of the changed items remained essentially the same in their primary content; the alterations related mostly to increasing the severity of the symptoms. This was done to decrease the number of people endorsing particular items, in the hope that the MCMI-III would be more selective in suggesting pathology. In addition, the items per scale were reduced by half, and the number of keyings was reduced from 953 on the MCMI-II to only 440 for the MCMI-III. The possible ratings per item were reduced from 1, 2, or 3 to either 1 or 2. The resulting 28 scales are divided into the categories shown in Table 8.1. Optimal BR cutoff scores were derived from a standardization sample of 1,079 clinical patients who had come from a diversity of backgrounds and treatment settings.

Theoretical Considerations

The development of the three versions of the MCMI has been partially guided by Millon's theories of personality. One of his core principles is the use of the polarities of pleasure-pain, active-passive, and self-other (R. Davis, 1999; Millon & Davis, 1996; Strack, 1999). These can be related to the fundamental evolutionary tasks of each person in that they must struggle to exist/survive (pleasure-pain), use various efforts to adapt to their environment or adapt their environment to themselves (passive-active), and invest in other people as well as themselves (other-self). Each of these polarities can be used to describe differences in personality organization for normal persons as well as those with personality disorders. For example, normal levels of functioning can occur on the active-passive dimension but, when these are exaggerated, they become dysfunctional. Thus, schizoid and avoidant personality disorders are extreme in the direction of passivity. In the self-other dimension, dependent and histrionic personalities are highly oriented toward others, whereas the narcissistic personality is extremely self-oriented. Many of the personality styles can be simultaneously portrayed on each of the three polarities. For example, the histrionic style is quite active and is both other (dependent) and pleasure-oriented. In some cases, the person is ambivalent on one or more of these dimensions, thereby resembling a person with a passive-aggressive style who is overtly passive and compliant but covertly expresses conflict and anger. Considerably more detail on these polarities, along with other aspects of personality disorders, can be found in Millon and Davis's (1996) *Disorders of Personality: DSM-IV and Beyond.*

Another important point relates to both the test's development and its implications for interpretation: The personality styles are not mutually exclusive. For example, a person with an antisocial style might be frequently uncomfortable with underlying anger and antisocial impulses and thus express them in passive-aggressive modes. This overlap also explains why the diagnosis of personality disorders has been plagued with poor interdiagnostician agreement (poor discriminant validity, R. F. Bornstein, 1998). The expected overlap among characteristics is one reason that the test developers were not overly concerned that many of the scales were highly correlated. Also, the overlap that was present seemed to occur in theoretically consistent patterns. From a practical perspective, this means that combinations of scale elevations can be used to give added meaning to each other. For example, a high score on the Antisocial scale, in combination with an elevation on the Sadistic scale, clearly suggests that the person will act out his or her antisocial feelings in a predictable and potentially dangerous manner. This activity would have very clear implications for case management and treatment planning.

Further, scale elevations should always be placed into the context of the person's life. A high score is not diagnostic of a personality disorder in and of itself. If a person can find an appropriate niche where the expression of his or her personality style is not dysfunctional, that person should not be considered "disordered." Thus, the distinction between a personality "style" and an actual personality "disorder" should be stressed. For example, a salesman with a narcissistic antisocial style might be able to optimize these traits in a way that makes him quite occupationally successful. The diagnostic criteria for personality disorders specifically state that there must be an enduring pattern leading to "clinically significant distress or impairment in social, occupational, or other important areas of functioning" (American Psychiatric Association, 1994, p. 630). If there is no or little distress or impairment, a personality disorder should not be diagnosed. This point is particularly crucial for the Compulsive (7), Histrionic (4), and possibly Narcissistic (5) scales because evidence is accumulating that these scales may not be measuring significant levels of pathology (Craig, 1999). Thus, it is crucial for clinicians to determine whether a given personality style suggested by the MCMI has actually led to distress and/or impairment.

Finally, the different categories of scales (Clinical Personality Patterns, Severe Personality Pathology, Clinical Syndrome, Severe Syndrome) are conceptually and clinically related (see Table 8.1). The first two categories relate to Axis II diagnoses but are separated to designate the greater levels of severity for the schizotypal, borderline, and paranoid conditions. As was previously pointed out, however, any of the personality styles are not disorders unless there is distress and impairment. The second categories are intended to measure the type and level of distress and thus relate more to Axis I levels of diagnoses. They represent the expression of personality styles that are not working well for the person. For example, if the narcissistic antisocial salesman mentioned previously tries to act toward his spouse as he does toward his business contacts, she may file for divorce. His means of coping with this outcome might be abuse of alcohol. In contrast, an individual with a dependent avoidant style who is undergoing a divorce would be likely to respond with a major depression. This difference underlies the essential interrelationship between Axis I and Axis II diagnoses. It also points out that the MCMI can help to establish the presence of an Axis II diagnosis by noting the type and degree of distress and impairment as expressed in elevations on the scales in the Clinical Syndrome and Severe Syndrome categories.

RELIABILITY AND VALIDITY

Reliability and validity studies on the MCMI indicate that it is generally a well-constructed psychometric instrument. Measures of internal consistency have been particularly strong. For the MCMI-III, alpha coefficients exceed .80 for 20 of the 26 scales, with a high of .90 for the Major Depression scale and a low of .66 for Compulsive (Millon, 1994, 1997). Test-retest reliabilities have been moderate to high. The MCMI-III manual reports that over a 5- to 14-day interval, test-retest reliability had a median of .91 (the high was .96 for Somatoform and the low was .82 for Debasement). Craig (1999) has summarized three data sets on test reliabilities ranging from 5 days to 6 months by stating that the median reliability was .78 for the Personality scales and .80 for the Clinical Syndrome scales. Much longer term test-retest reliabilities spanning 4 years ranged from a high of .73 for Passive-Aggressive to a low for Dependent of .59 (Lenzenweger, 1999). This is roughly equivalent to other stable dimensions of personality.

Because the personality scales theoretically represent enduring, ingrained characteristics, they should have greater stability than the clinical scales, which are based on more changeable symptomatic patterns. In some cases, this has been found to be true; in others, little difference has been found. Studies on the MCMI-I have indicated the theoretically expected higher stability for the personality scales as opposed to the clinical scales (Piersma, 1986). In contrast, the Craig (1999) summary found very little difference between the mean personality and clinical scales, despite an extended retesting interval. Similarly, the MCMI-III manual reported a mean of .89 for the personality scales and a slightly greater mean of .91 for the clinical scales. This suggests that the original MCMI may have had the theoretically higher temporal stability for the personality scales versus the clinical scales, but later versions have roughly equivalent temporal stabilities between the two categories of scales.

One central issue when evaluating the validity of the MCMI is the extent to which validity studies on previous versions can be generalized to the newer versions. With appropriate caution, some transferability is probable because the correlations between the MCMI-II and MCMI-III scales are moderately high. Specifically, the correlations range from a high of .94 for Debasement to a low of .59 for Dependent, with 12 of the 25 scale comparisons above .70. Comparisons between the Depressive and Posttraumatic Stress Disorder scales could not be made because they were uniquely developed for the MCMI-III. However, Marlowe, Festinger, and Kirby (1998) found much lower correlations and little comparability for code types. This should be balanced with their conclusion that the MCMI-III did provide comparable clinical information. It should also be noted that they used a relatively small sample of persons who had a quite specific disorder (cocaine dependence). It would thus be important for future studies to determine the extent to which the MCMI-III is actually comparable to the MCMI-II. Some researchers consider the MCMI-III to be sufficiently different to be considered a separate instrument (Marlowe et al., 1998; R. Rogers, Salekin, & Sewell, 1999). Rogers et al. (1999) have even suggested that, in contexts requiring maximum accountability (forensic), it might be preferable to use the MCMI-II until the validity of the MCMI-III becomes better documented. Given these considerations, the following selective and representative overview of validity draws on material from both the more recent MCMI-III and the MCMI-II.

Factor analysis of the MCMI-II has generally supported the organization of the scales. The most extensive published factor analysis involved 769 cases and resulted in

an eight-factor solution (Millon, 1987). The largest factor accounted for 31% of the variance, was related to general Maladjustment, and involved depressed affect, impaired interpersonal relationships, low self-esteem, and unusual cognition and self-behavior. The next two largest factors were Acting Out/Self-Indulgent (13% of the variance) and Anxious and Depressed Somatization (8% of the variance). The final factors, listed according to progressively decreasing proportions of the variance, were Compulsively Defended/Delusional Paranoid, Submissive/Aggressive Sadistic, Addictive Disorders, Psychoticism, and Self and Other Conflictual/Erratic Emotionality. More than 20 factor-analytic studies have been performed on the various MCMI versions, and these have generally supported the keying of the items (Retzlaff, Lorr, & Hyer, 1989) as well as the clustering of the factors around Millon's conceptualization of psychopathology (Choca, Retzlaff, Strack, Mouton, & Van Denburg for 1996; Choca & Van Denburg, 1997; McCann, 1991). Craig and Bivens (1998) performed a factor analysis on the MCMI-III using 444 outpatients and found three factors they labeled General Maladjustment, Paranoid Behavior/Thinking with Detached Acting Out, and Antisocial Acting Out.

A variety of correlations have been made between the MCMI and various related instruments, including the Beck Depression Inventory, General Behavior Inventory, Michigan Alcoholism Screening Test, State-Trait Anxiety Inventory, Symptom Checklist-90, and the MMPI (Millon, 1994, 1997). These are reported in detail in the MCMI-III manual. Representative findings include expected correlations between the Beck Depression Inventory and the MCMI-III Major Depression (.74) and Dysthymia (.71) scales. Similarly, high correlations were found between the MMPI-2 Depression scale and the MCMI-III Depression (.71) and Dysthymia (.68) scales. As would be expected, negative correlations were found between the Beck Depression Inventory and MCMI-III scales related to denying pathology (Histrionic, –49; Narcissistic, –40; and Compulsive, –30). An additional representative finding was a .55 correlation between the MCMI-III Somatoform scale and the Symptom Checklist-90-Revised scale for Somatization. One puzzling finding was a low correlation of .29 between the MCMI-III Paranoid scale and the MMPI-2 Paranoia scale. Similar surprising results were moderate correlations between the MMPI-2 Psychopathic deviate scale and the MCMI-III scales for Self-Defeating (.45), Schizotypal (.43), and Depressive (.41). For the most part, however, correlations between the MCMI and external criterion instruments have been in the expected direction.

One of the important and relatively unique contributions of the MCMI has been the development and availability of data on its diagnostic efficiency. This is usually calculated by designating BR scale scores of 75 and/or 85 as *test positives* and comparing these with clinician ratings of whether the characteristics predicted by the scale scores actually matched these clinician ratings. In some settings, however, it is important to take into account the frequency by which a disorder occurs in that setting (BR of the disorder). For a test to be effective, it must diagnose a disorder more accurately than the chance occurrence as determined by the BR. For example, forensic and/or substance abuse treatment facilities usually have high numbers of persons with antisocial personality styles. In these cases, calculation of the *positive predictive power* of the MCMI for the particular setting is recommended. Essentially, positive predictive power is a calculation of the probability that a test score accurately

indicates the presence of a characteristic or diagnosis based on some other measure such as clinical ratings. Such a calculation involves a formula (Gibertini, Brandenberg, & Retzlaff, 1986; see Millon, 1994, pp. 41–43) in which prevalence rates must be inserted (derived from knowledge regarding a specific client population) along with sensitivity and specificity data (available in the MCMI-III manual). Such a calculation provides practitioners with an estimate of the extent to which the instrument performs beyond merely BR levels. For example, if the prevalence or BR of antisocial personalities is .25 but the positive predictive power of the MCMI is .76, the difference (.76 – .25) of .56 indicates that the incremental validity of the instrument is .56 above merely BR (prevalence) or chance predictions. This emphasis on levels of certainty, with its implications for actual clinical decision making, is one of the strong features of the MCMI.

Calculations of the positive predictive power of the MCMI-II indicated good predictive power ranging between .30 and .80 (Millon, 1987). This was supported by R. Rogers et al. (1999), who pooled existing data on the convergent/divergent validity of the MCMI-II scales and found good support for Avoidant (2A), Schizotypal (S), and Borderline (C); and moderate support for Schizoid (1), Dependent (3), Histrionic (5), Antisocial (6A), Aggressive (6B), Negativistic (8A), Self-Defeating (8B), and Paranoid (P). Little support was found for Compulsive (7%). Positive predictive power for the MCMI-III Axis II scales indicated that the highest accuracy was found for the Dependent (81%), Paranoid (79%), and Compulsive (79%) scales (Millon, 1997). In contrast, relatively low positive predictive power was found for the Masochistic (30%), Negativistic (39%), and Depressive (49%) scales. A similar study by R. Davis, Wenger, and Guzman (1997) found that the highest positive predictive power was found for Dependent (.81), Paranoid (.78), and Compulsive (.79), whereas the lowest was found for Masochistic (.30), Negativistic (.39), and Depressive (.49). The low predictive power for Masochistic and Negativistic may, in part, be attributed to the fact that Masochistic/Self-Defeating was entirely deleted from the *DSM-IV* and Negativistic and Depressive were relegated to the appendix. This meant the *DSM-IV* considered them to be poor diagnoses in the first place; and it also meant that, for the masochistic diagnosis, clinicians did not have the assistance of *DSM-IV* guidelines. Both studies concluded that comparisons of the three generations of the MCMI generally indicate progressive increases in its psychometric characteristics in general and, more specifically, in its diagnostic accuracy.

ASSETS AND LIMITATIONS

The strategy of developing the MCMI has been commendable and innovative. The history and development section outlines how this has involved a combination of theoretical-conceptual, internal-structural, and external criterion procedures. Each of the procedures has progressed in a stepwise manner; only those items that survived the previous steps were retained. The result has been an instrument that adheres closely to theory, demonstrates good reliability, and has shown excellent promise regarding internal and external validity. The use of BR scores has been a noteworthy innovation and has probably resulted in increases in diagnostic accuracy. However, difficulties have been noted related to the extensive item overlap and low level of interdiagnostician

agreement among clinicians using methods such as structured interviews and the MMPI. The scale abbreviations are also "user unfriendly."

As pointed out previously, the MCMI is a relatively time-efficient test that potentially produces a wide range of information. Of central importance, this information focuses not only on clinical symptomatology (Axis I), but also on the more enduring and potentially more problematic personality disorders (Axis II). These personality disorders can frequently be overlooked. Practitioners might overlook them because (a) the client is more likely to express concern over more overt symptoms, and (b) personality styles are often more hidden and must be inferred. Clients themselves may be unaware of personality styles that have become so automatic that it is difficult to recognize them. They can feel the emotional pain of symptoms, but are rarely aware of the recurring patterns of behaviors and cognitions that frequently are at the core of the development and maintenance of these symptoms. In addition to knowing the client patterns that lead to symptoms, considerable literature supports the usefulness of knowing a client's status related to personality disorder. For example, a personality disorder diagnosis suggests that the client is at risk for interpersonal difficulties; these difficulties may complicate the therapeutic relationship and alter the course of Axis I disorders (R. F. Bornstein, 1998). Turkat (1990) has estimated that 50% of clients seeking psychotherapy meet the criteria for a personality disorder. Thus, the MCMI inhabits a crucial niche in objective assessment because it has been designed to better understand personality dysfunction.

Despite the assets of the MCMI, there are a number of inherent difficulties in the assessment of personality disorders. One central issue is that there is no "benchmark" or "gold standard" with which to compare the MCMI assessments. Individual clinicians relying on interview information generally have low interdiagnostician agreement (median kappa = .25; J. Perry, 1992). This is sufficiently low enough that it would be unacceptable in any other area of psychological research. Similarly, formal instruments such as the MMPI, MCMI, and structured interviews have shown little agreement (H. Miller, Streiner, & Parkinson, 1992; Streiner & Miller, 1990), which makes it difficult to judge the "true" accuracy of MCMI personality disorder assessment. Several attempts to deal with this have been made. R. F. Bornstein (1998) urges diagnosticians to simply accept that "our ability to describe different personality disorders has outstripped our ability to diagnose them accurately in real-world clinical settings" (p. 334). His solution is to decide whether a client has a personality disorder (which does have very good interdiagnostician agreement; Loranger et al., 1995) and then rate the intensity and level of impairment of various personality characteristics. For example, it might be decided that a client has a personality disorder with dependent (moderate intensity, low impairment levels) and histrionic (low intensity, moderate impairment) features. In contrast, Westen and Shedler (1999a, 1999b) have pointed out that the actual arena of diagnosis by clinicians occurs when they infer personality characteristics from client narratives and match the extent to which these inferences match prototypical conceptions of personality disorders. By rank-ordering composite descriptions of prototypical personality disorders, they developed the Shedler-Westen Assessment Procedure-200 (SWAP-200). Clients can then rate a particular client and note the extent to which that client meets the ideal descriptions in the SWAP-200. Initial studies have shown the scale has high alpha (above .90 for 14 of 15 diagnoses), good convergent/divergent validity, and supportive factor analysis (Westen & Shedler, 1999a,

1999b). Future research might compare these procedures with MCMI-III data and especially determine the extent to which the MCMI-III might demonstrate incremental validity beyond them.

A related issue is that some of the diagnostic criteria incorporated into the MCMI items are closely tied to the *DSM* criteria, whereas others are more closely linked to Millon's theories (see R. Davis, 1999). In some cases, these criteria are similar; in others, the criteria are different. This has led to some controversy regarding the relative advantages and disadvantages of having different criteria (P. Flynn, McCann, & Fairbank, 1995). One disadvantage is that, in many cases, the MCMI should not be considered a *DSM* measure even though the titles of the scales may lead practitioners to think that it is (Wetzler & Marlowe, 1992; Widiger, Williams, Spitzer, & Francis, 1985). However, this may actually be an advantage for some of the scales/disorders because the *DSM* criteria have been criticized as being both insufficiently related to theory and clearly inadequate in some areas. Because the MCMI has not strictly adhered to the *DSM* criteria, it can work to remedy some of the *DSM*'s perceived inadequacies. For example, the *DSM-III-R/DSM-IV* diagnosis for antisocial personality has attained high interrater reliability but has done so by sticking closely to clear behavioral criteria primarily related to overt acts against society. The more intangible but crucial issue of poor conscience development has not been sufficiently addressed, which has led to accusations that the *DSM* criteria relate more to a "criminal" disorder than to a "personality" disorder. The theory behind the MCMI antisocial personality disorder items stresses both overt behaviors and the relative lack of conscience, and this conceptualization is reflected in the item content.

One further issue relevant to the diagnosis of personality disorders is the difficulty in distinguishing state and trait. Theoretically, Axis I disorders relate primarily to states, and Axis II characteristics relate to traits. In reality, they are highly interdependent and it is often difficult to separate them. State (clinical) MCMI elevations seem to be closely related to scores on trait (personality) scales. For example, J. Reich and Noyes (1987) found a 50% decrease in MCMI personality disorder prevalence estimates when the MCMI was given during the recovery phase as opposed to measures during the acute phase. Elevations in MCMI-II personality scales have also been demonstrated to increase the more state-related MMPI-2 F (and other validity) scales (Grillo, Brown, Hilsabeck, Price, & Lees-Haley, 1994). Given this state/trait distinction, it would be predicted that the trait/personality scales would be more stable than the clinical/state scales; yet, in many cases, this has not been demonstrated to be so. A number of sources, including the MCMI-III manual (Craig, 1999; Millon, 1994, 1997), have demonstrated little difference in the test-retest reliabilities between the two categories of scales. Collectively, these observations indicate that state and trait measures are quite interdependent. To account for this, Millon has developed, for some of the MCMI-II and MCMI-III scales, a number of adjustments that work similarly to the *K* correction on the MMPI. Also, as with the MMPI *K* correction, it is unclear and controversial as to how effectively they achieve their purpose.

Because of the MCMI's reliance on the *DSM,* efforts have been made to incorporate changes that parallel the ongoing developments of the *DSM.* This has the advantage of keeping current with changing diagnostic criteria, but it has also meant that the MCMI has been relatively frequently revised (Millon, 1977, 1987, 1994). In contrast, the MMPI and CPI have been through far fewer revisions. As a result, the

MCMI (and particularly the MCMI-II) has not been in an unchanged version long enough for it to be fully evaluated. With the introduction of the MCMI-III in 1994, researchers and practitioners must now wait until sufficient research has been performed to decide such crucial questions as whether the MCMI-III has sufficient validity, when (or whether) to continue with the previous MCMI-II, or whether the interpretations developed for the MCMI-I and MCMI-II should now be used with the MCMI-III.

An important consideration is whether the MCMI measures actual personality "disorders" or, rather, personality "style." As indicated previously, the MCMI-III measures of histrionic, compulsive, and, possibly, narcissistic traits do not seem to be measuring actual disorders but more styles (Craig, 1999). Choca and Van Denburg (1997) prefer to think of the various scales as referring to personality "style" because the inference to disorder requires more information than can realistically be found in scale elevations. Persons with certain personality styles may have been able to find an occupational and/or interpersonal niche that allows them to function adequately. For example, a Schizoid or Avoidant personality may work quite well as a night watchperson. Thus, the inference from style to disorder must be made by the individual practitioner and not by the test. Practitioners who look for the test to include actual diagnosis are overextending its use beyond realistic expectations.

A further issue with the MCMI has been extensive *item overlap.* The original MCMI (MCMI-I) had its 175 items arranged on 733 different keyings, and the MCMI-II had an even greater 953 keyings. Thus, because many of the items were used to score numerous scales, there were frequently high scale correlations. For example, the MCMI-I's Borderline and Dysthymia scales shared 65% of their items and were highly correlated (.95). Given these characteristics, practitioners might be justifiably concerned that some of the scales were measuring constructs that were too similar and therefore redundant. The defense of the high scale intercorrelation has been that many of the constructs are theoretically and clinically similar, and the similarity would, therefore, be psychometrically reflected in many high scale correlations. Practitioners would then need to look at the relationships between the different scale elevations as a means of "fine tuning" their interpretations. For example, Avoidant and Schizoid personalities are similar in their passivity and interpersonal distance. Clinical lore suggests that many persons initially believed to be Schizoids appear more similar to Avoidants as more information is obtained from them. Given the theory and intent of the MCMI, the two scales measuring these styles would be expected to have similar items and to be elevated. Notwithstanding this defense, the recently revised MCMI-III has attempted to reduce the item scale overlap (and resulting intercorrelations) by reducing the number of items per scale, providing item weightings depending on their relative importance for a scale, and reducing the number of keyings to 440 (Millon, 1994). This seems to have been successful in that none of the interscale correlations reported in the manual were above .90 and only three were .80 (the rest were lower). This suggests that the MCMI-III scales, compared to the previous versions, are measuring somewhat more independent domains.

When interpreting the MCMI, it is sometimes difficult to know where the interpretive information was derived. It is based on a combination of theory and empirical relationships determined specifically through validity studies of the MCMI itself. Each of these two interpretive sources has developed over a number of years, during which three versions of the MCMI have appeared. It is often difficult to know whether the

interpretations have been empirically versus theory-based or whether they have been derived from validity studies done on previous versions of the MCMI. If done on previous versions, it can be rightfully argued that most of the interpretations can be transferred from these earlier versions because there has been continuity in theory and scale development. This is particularly reflected in the moderate-to-high correlations between the new and the older scales. However, practitioners must struggle with which of the interpretations have been empirically versus conceptually derived as well as which are obsolete versus still current. This problem is relevant for the MCMI as well as other similar instruments (i.e., MMPI, CPI), and it highlights the importance of clinicians' working with the test results and integrating them with additional sources. Millon (1992) summarizes that the quality of the interpretive information is dependent on "the overall validity of the inventory, the adequacy of the theory that provides the logic underlying the separate scales, the skill of the clinician, and the interpreter's experience with relevant populations" (p. 424).

A criticism related to this issue is that the MCMI overdiagnoses and overpathologizes (P. Flynn et al., 1995). For example, Wetzler (1990) has noted that MCMI-related diagnoses of personality disorder were 60% higher than diagnoses based on structured interviews. One of the reasons for at least the potential for overdiagnosis among practitioners is the possibly misleading name of the personality scales. They create the external appearance of clear *DSM* diagnostic categories when they are probably best conceptualized as styles that may or may not reflect an actual disorder. This does not mean that the MCMI cannot be extremely useful in diagnosing personality disorders; but it should be more accurately perceived as placing the practitioner in the correct domain or coming half (or more) of the way toward diagnosis. A further problem is that the MCMI does not perform well on normal or only mildly disturbed populations. Such persons might have moderate elevations that are still below the BR 75 cutoff, but practitioners might be tempted to interpret these "elevations." This confusion is complicated further if correlations derived from the Modifying Indices "bump up" the personality or clinical scales into the interpretable range. Unfortunately, the MCMI National Computer System computer interpretations tend to both reinforce interpretations of moderate "elevations" and suggest that *DSM* diagnoses can be made based on MCMI scores. Thus, the MCMI should be used only with clinical populations. A related difficulty is that the scales and their related interpretations tend to emphasize a client's deficiencies without balancing these out with the client's strengths. The result is likely to be an overly negative description of a client's functioning. This occurs despite the fact that many aspects of personality styles might be quite adaptive; for example, the easy sociability of the histrionic style or the adaptability and empathy associated with many persons with depressive styles (see R. F. Bornstein, 1998; Westen & Shedler, 1999a, 1999b). In addition to overpathologizing, the MCMI-III has also been found to perform poorly when assessing persons with psychotic disorders (Craig, 1999).

INTERPRETATION PROCEDURE

Effective interpretation of the MCMI requires considerable sophistication and knowledge related to psychopathology in general and personality disorders in particular. At a minimum, practitioners should be familiar with issues related to personality disorders,

along with the *DSM-IV* criteria. Ideally, practitioners should also have read Millon and Davis's (1996) definitive *Disorders of Personality: DSM-IV and Beyond,* worked with clients with personality disorders, and administered the MCMI to a number of such clients. Clinicians should also be aware of the previously outlined assets and limitations of the MCMI so that they can most appropriately work with the data. In particular, the MCMI does not provide *DSM-IV* diagnosis; it should be used only with clinical populations; it is not particularly helpful in assessing a person's strengths; and there is a possibility that it might overdiagnose personality disorders and be overinterpreted by clinicians.

One consideration in interpreting the MCMI is the possible influence of gender, age, and ethnicity. Gender influences have been minimized by using separate norms for scoring the profiles of males and females. The gender differences that have emerged on the MCMI are also consistent with prevalence rate estimates. For example, the greater rate of antisocial personalities among males is reflected in the BR scores, which take this greater prevalence rate into account. Some differences between European American and African American psychiatric patients have been found on nine of the 20 MCMI-II scales. African Americans scored especially higher on Antisocial, Narcissistic, Paranoid, Hypomania, and Drug Abuse scales (Choca, Shanley, Peterson, & Van Denburg, 1990). As with the MMPI, ethnic differences have been noted, but the meaning attributed to these is less clear. For example, the greater elevations on these MCMI-II scales may mean that these scores are accurate representations of the more difficult circumstances many African Americans encounter. In contrast, the higher scores might be distortions based on error in the specific scales. This controversy should alert practitioners to be cautious and conservative when making interpretations for African American or other ethnic groups. Finally, there do seem to be age-related differences on the MCMI-II: Older persons score higher on Dependent but lower on Compulsive and Borderline (Choca, Van Denburg, Bratu, & Meagher, 1995). This means that interpretations among older persons should take these age-related variables into account.

The set of procedures outlined in the next section is recommended for interpreting the MCMI. The discussion of the various scales and codes represents an integration and summary of current research as well as material included in the MCMI-III manual (Millon, 1994) and interpretive guides developed by Choca and Van Denburg (1997), Craig (2001), and Jankowski (2002). The subsections related to treatment planning have summarized material from Dorr (1999); Goncalves, Woodward, and Millon (1994); Retzlaff (1995); Retzlaff and Dunn (2003); and Millon and Davis (1996). Each of the 28 MCMI-III scales is discussed in relation to interpretation, possible interaction with other scales, and implications for treatment planning.

The formal elaboration and separate listing of two- and three-point code types are not discussed, for several reasons. First, research on the MCMI does not have the well-developed code-type validity literature found for the MMPI. Instead, many of the MCMI code-type descriptions are based on a conceptual integration of the implications of clusters of scale elevations. In many ways, this is a task that individual practitioners can do themselves by rationally considering the meanings of associated scale elevations. For example, an elevation on Antisocial, combined with a corresponding elevation on Aggressive (Sadistic), would clearly indicate the abusive, combative, and impersonal expression of the person's antisocial tendencies. Second, given that there

are fully 28 MCMI-III scales, the total number of possible code types is both unwieldy to list and unrealistic to fully research. However, a short subsection (Frequent Code Types) under most of the scale descriptions does briefly describe the meanings attached to some of the more important associated scale elevations. Readers are encouraged to read these descriptions and to expand on their meanings by reading the longer interpretive descriptions for the entire associated scale.

1. Determine Profile Validity

Before interpreting the personality and clinical scales, practitioners must be assured that the client has not over- or underreported symptoms or responded in a random manner. The profile validity can be assessed by noting the pattern of scores on the Modifying Indices (validity indicators):

Random responding is suggested by scores of one or more on the three items of the MCMI-III Validity scale ("True" on items 65, 110, and 157).

Underreporting of difficulties on the MCMI-III is suggested by low scores (raw score less than 34) on Disclosure (X) and Debasement (Z) and a high score (BR over 75) on Desirability (Y; an "arrow" profile on the Modifying Indices). However, it is sometimes difficult to differentiate persons who are faking good (underreporting) from those who actually have the positive qualities of being cooperative, self-confident, and conscientious. The client's history is often the best tool for making this distinction.

Fake bad profiles are suggested by a high score (raw score above 178) on Disclosure (X) and a high score (BR above 75) on Debasement (Z; a "valley" profile on the Modifying Indices). With moderate elevations, this might be a "cry for help"; but with progressively higher scores (BR above 85), the likelihood of an invalid profile is increased.

It should be noted that BR adjustments for certain scales have been made in an effort to increase MCMI-III profile validity. The adjustments serve as correction scores in much the same way as the K correction serves for the MMPI. These adjustments are part of the standard scoring and involve adjustments for Disclosure (if either high or low), Anxiety/Depression, Inpatient, and Denial/Complaint.

2. Interpret the Personality Disorder Scales

Retzlaff (1995) recommends that, when interpreting the personality disorder scales, practitioners should first check to see whether any of the Severe Personality Disorder scales are elevated. If so, this strongly suggests that one or more of the Clinical Personality Pattern scales will also be elevated. However, the high scale(s) on the Severe Personality Disorder section should take precedence over equivalently elevated scales on the Clinical Personality Pattern scales. The Clinical Personality Pattern scales then serve to color or elaborate on the elevation(s) on the Severe Personality Disorder scale(s). The primary focus for diagnosis, then, should be to rely on the Severe Personality Disorder elevation unless elevations on other categories of scales were extremely elevated compared to the Severe Personality Disorder scales. When that occurs, the

extremely elevated scales would take on greater interpretive meaning compared to the more moderately elevated Severe Personality Disorder scale(s). (Interpretive descriptions of each of these scales can be found in the next section.) If there are no elevations on the Severe Personality Disorder scales, practitioners should interpret any elevations on the Clinical Personality Pattern scales.

The interpretive sections under the personality disorder scales are divided into general interpretive descriptions, frequent code types, and treatment implications. Often, the descriptors are fairly severe and negative. Interpreters need to determine whether these apply to the individual client, based on how high the scale is elevated, implications of associated scale elevations, and additional data available on the client. For example, some of the descriptors might need to be "softened" if they are in the marginally elevated range (BR 75 to 80). More severe interpretations might be appropriate for extremely elevated scores or if the elevations are from either the Severe Personality Pathology or Severe Syndromes categories. Millon (1994) specifies that scores in the 75 to 84 range indicate the syndrome or pattern is present, whereas scores of 85 or above indicate that it is prominent. The general rule, then, is: The higher the elevation, the more likely that the interpretive descriptions are accurate. Another consideration is the height of elevated scales relative to other elevated scales. If they are approximately the same height, they should be given equal interpretive weight. On the other hand, if there are 20 or more BR points between scales, the lower scale's influence is likely to be so subtle that it can be minimized or even ignored.

The general interpretive descriptions for the Clinical Personality Pattern scales (but not the Severe Personality Disorder) also include paragraphs on possible strengths or positive descriptions. These provide a means of partially balancing the primarily negative descriptions associated with scale elevations. The Frequent Code Types subsection gives a brief description of the meanings attached to frequently associated elevations. Any code types that have been described previously refer the reader back to these earlier descriptions.

3. Interpret Clinical Syndrome Scales

Similar to step 2, Retzlaff (1995) recommends that precedence be given to interpreting any elevations on the Severe Clinical Syndrome scales. Sometimes, all or most of these scales are elevated, which should not be considered contradictory; rather, these elevations can be used to complement one another. Any elevations on the Severe Clinical Syndrome scales are usually accompanied by complementary elevations on the Basic Clinical Syndrome scales as well as the Personality Disorder scales. For example, an elevation on the Severe Clinical Syndrome scale of Major Depression might also have corresponding elevations on Drug Dependence, Anxiety, and Avoidant. Interpretations would center on depression but would also include fear of interpersonal involvement, anxiety, and the distinct likelihood that the person is using alcohol as a means of coping with these difficulties. Another example might be a person with an elevation on Anxiety but with a corresponding elevation on Avoidant and Dependent, which suggests he or she is experiencing the anxiety because of conflict between wanting to be accepted and cared for by others, and being terrified of criticism and humiliation. In contrast, another person with an elevation on Anxiety but with a corresponding elevation on Narcissistic is most likely experiencing anxiety because of significant challenges to his or

her self-inflated sense of importance and superiority. This careful interplay between the scales is crucial for accurate and effective profile interpretation.

One of the unique features of the MCMI is that it is an objective test that measures personality styles/patterns relevant to Axis II disorders. The sections describing each of these scales include subsections on frequent code types (including possible relations with Clinical Syndrome scales) and treatment implications. In contrast, the Clinical Syndrome scale descriptions include only descriptions of the scales, without material on frequent code types or treatment implications, partially because relevant relations with the personality scales are mentioned in the previous section. In addition, there is already a well-developed clinical literature (and extensive time is spent in most training programs) on treating these clinical syndromes (anxiety, depression, etc.). Practitioners who wish to develop detailed treatment procedures for difficulties measured by the Clinical Syndrome scales can consult resources such as Barlow's (2001) *Clinical Handbook of Psychological Disorders* or Kaplan and Sadock's (2001) *Kaplan & Sadock's Pocket Handbook of Clinical Psychiatry.*

4. Review Noteworthy Responses (Critical Items)

The MCMI-III manual (Millon, 1997) has listed a series of Noteworthy Responses in Appendix M (pp. 181–182). These are organized around the topics of Health Preoccupation, Interpersonal Alienation, Emotional Dyscontrol, Self-Destructive Potential, Childhood Abuse, and Eating Disorder(s). Similar to the MMPI critical items, the MCMI's Noteworthy Responses are not so much formal scales as they are rationally categorized items that might be important for a clinician to more fully understand. Accordingly, they can be used to organize a semistructured interview around relevant responses. They can also be selectively inserted into a psychological report to provide a more concrete qualitative portrayal of the client's attitudes, affect, and behavior.

5. Provide Diagnostic Impressions

Given the interpretive descriptions of a client's profile (steps 2, 3, and 4), along with any other relevant information, clinicians can formulate the most appropriate diagnosis.

6. Elaborate on Treatment Implications and Recommendations

The symptoms reported and reflected in elevations on the Clinical Syndrome scales (Anxiety, Depression, Substance Abuse, etc.) are those that are most problematic and thus should be targeted as high priorities. However, these also need to be understood in the context of the client's personality patterns and pathologies. Under each of the personality disorder scales, there are sections with relevant suggestions for treatment recommendations. These can be considered, along with other information, to expand on what would be the most appropriate interventions. Additional useful resources in this process are Chapter 14 of this book ("Psychological Assessment and Treatment Planning"), Millon and Davis's (1996) *Disorders of Personality: DSM-IV and Beyond,* and Retzlaff's (1995) *Tactical Psychotherapy of the Personality Disorders: An MCMI-III-Based Approach.*

MODIFYING INDICES (VALIDITY SCALES)

The MCMI Modifying Indices are adequate at detecting random responding, fake bad, and fake good profiles. However, the detection rate appears lower than for the MMPI (Bagby, Gillis, Toner, & Goldberg, 1991), and, as with the MMPI, fake bad profiles are more accurately detected than fake good (defensive) profiles (Fals-Stewart, 1995; Millon, 1987). Using the decision rules for fake bad profiles, the rate of accurate detection runs between 48% and 92% (Bagby, Gillis, Toner, et al., 1991; Retzlaff, Sheehan, & Fiel, 1991). However, for severely disturbed clients, high scores on fake bad indices may be more indicative of high distress and a "cry for help" than an invalid profile (Wetzler & Marlowe, 1990). In contrast to the generally good detection rate for fake bad profiles, persons faking good (defensively) are likely to be detected approximately 50% of the time (Retzlaff et al., 1991), and clients underreporting their substance abuse seem particularly good at avoiding detection (Fals-Stewart, 1995). This means that the MCMI should be used with extreme caution in situations in which individuals might be likely to underreport their psychopathology.

The most useful tool in making these decisions related to validity is a careful consideration of the client's past and current level of functioning. Specifically, a person who may look as though he or she is faking bad, but whose history reveals a person who is dysfunctional, may be merely expressing distress. In contrast, a relatively highly functioning person with the same scores on Modifying Indices is much more likely to be faking bad. Conversely, a person with a potentially fake good profile but who also has a high level of functioning may be merely expressing actual confidence, assertiveness, and high self-esteem. In contrast, a person with a similar profile but a history of interpersonal, legal, and/or psychiatric history is much more likely to be underreporting psychopathology.

Validity Index (Scale V)

The MCMI-III Validity Index is composed of three items (numbers 65, 110, 157) that, if endorsed as true, indicate absurd responses. As a result, endorsement of these items strongly suggests that a person has responded randomly. The manual states that one true response should be interpreted as indicating a profile of "questionable validity," and two or more endorsements can clearly be interpreted as an invalid profile. Presumably, the "questionable validity" option is given to suggest that the profile *may* still be valid in the event that the client has misread or randomly responded to only a few items (including one of the three on the Validity Index). This allows the possibility that most of the items were still responded to accurately. In contrast, Bagby, Gillis, and Rogers (1991) recommend that even one endorsed item be used to indicate an invalid profile. One caution: If a person did respond randomly, there is still a statistical chance that he or she may have gotten "lucky" and none of the three items were answered in a true direction, in which case detection on the Validity Index would be avoided. In addition, a person wishing to consciously fake responses would be able to notice the absurdity of answering "true" to any of the Validity Index questions and would answer them in such a way that they would not endorse scorable responses on the Validity Index.

Disclosure Index (X)

The Disclosure Index was designed to measure whether a client's responses were open and revealing as opposed to defensive and secretive. If the MCMI-III raw score on the Disclosure Index is below 34, it most likely indicates a defensive underreporting of psychopathology. It may also mean that the person did not read or understand the questions correctly. A further interpretation is that the client is hesitant, reserved, and overconcerned with seeking social approval. However, low Disclosure Index scores on the MCMI-II were not found to be particularly sensitive because subjects requested to "fake good" still produced generally acceptable Disclosure Index scores (Retzlaff et al., 1991). This means that when clients do fake good extensively enough to produce a clearly low Disclosure Index, the profile can be considered invalid with a fair degree of certainty.

MCMI-III raw scores above 178 indicate that the individual has extensively exaggerated his or her symptoms. The reporting of symptoms would even exceed fairly disturbed psychiatric populations and, therefore, suggests an overreporting of symptoms. This is the only scale on the MCMI that is interpreted if either high or low. The other scales on the MCMI should be interpreted only if they are above the BR 75 cutoff.

Desirability Index (Y)

Similar to the Disclosure Index, the Desirability Index is also a measure of defensive responding. Scores above BR 75 indicate the individual has presented in a manner that is unusually moral, interpersonally attractive, extremely emotionally stable, highly gregarious, organized, and with a high respect for the rules of society. Progressively higher scores suggest that the person is concealing crucial details regarding psychological or interpersonal difficulties. However, this is not a particularly good scale and should be interpreted with caution.

Debasement Index (*Z*)

As the title of the scale suggests, the Debasement Index reflects the extent to which a person is describing himself or herself in negative, pathological terms. This might include feelings of being empty or angry, crying easily, having low self-esteem, possibly being self-destructive, and frequently feeling tense, guilty, and depressed. Thus, the Debasement Index measures characteristics opposite from those on the Desirability Index. They are rarely both elevated although, on occasion, someone who is unusually self-disclosing may have high scores on both. Scores above BR 85 indicate either a cry for help resulting from acute psychological distress or a fake bad profile. As with Desirability, this scale is not particularly effective and should be interpreted with caution.

CLINICAL PERSONALITY PATTERNS

Schizoid (Scale 1)

The core characteristic of persons with elevations on this scale is little or no interest in other people. Their lives are spent as loners. They are detached, impersonal, withdrawn,

unsociable, seclusive, passive, and distant, and have few, if any, friends. They rarely initiate conversation, are indifferent to other people, and rarely seek involvement with others. In family, work, or social situations, they prefer to have a peripheral role. As a result, they frequently function on the margins of society. They have little drive to have their needs met, experience few erotic attachments, express little warmth, and are often asexual. Rarely do they experience very much depth of feeling (pleasure, sadness, anger). They are largely indifferent to praise or criticism from others. Their interpersonal distance is not based on a defense stemming from fear of rejection but is, rather, their natural and most comfortable way of functioning. They also lack vitality and are unanimated and almost robotlike in their movements. When they do communicate with others, it is in a vague, distant, unfocused manner. Often, the direction of their conversation loses its focus and whatever information is conveyed is delivered in a circuitous manner. As a result, others are likely to see them as strange or "spacey." They have little self-awareness or insight into the implications of interpersonal relationships. If they are involved in a committed or intimate relationship, a frequent spousal complaint is that there is insufficient closeness, sharing, and understanding.

An asset of this personality style is that these persons typically do not become particularly disturbed by anything. Although they are not particularly involved with or interested in others, when they do interact, they are typically quite comfortable. Decision making is often easier for them because it is not complicated by emotional or interpersonal intricacies. They are also quite self-sufficient—they are comfortable with spending extensive periods of time alone and may have a rich fantasy life. Their hobbies typically involve activities that require only minimal contact with other people.

Frequent Code Types

Clinical scales that are likely to be elevated along with Schizoid are Anxiety and Thought Disorder. This reflects the sometimes obsessive thinking of the schizoid, along with the possibility that brief psychotic states might occur. Personality scales that are often elevated along with Schizoid are Avoidant, Passive-Aggressive (Negativistic), Dependent, and Compulsive. Each of these adds new variations onto the previous description. A corresponding elevation on Avoidant suggests that these persons are not only uninterested in and unskilled at interpersonal relationships, but also uncomfortable around others and fear rejection. However, behind their detachment may be a real desire to become involved. If Schizoid and Avoidant are both elevated, the possibility of problem use of alcohol should be investigated (check Alcohol and Drug Dependence). Elevations on the Passive-Aggressive (Negativistic) scale (along with Avoidant) underscore conflictual feelings and possible resentment toward the few interpersonal relationships they have. This resentment centers on a wish that someone would nurture and guide them (especially if Dependent is also elevated) along with fear that they might be rejected. This conflict results in their frequently being moody and nervous. An elevation on Dependent (along with Avoidant) indicates that they feel less important and capable than others. As a result, they are submissive, humble, and congenial as a means of seeking acceptance and being cared for. When Compulsive is elevated, these persons are disciplined, well organized, emotionally controlled, meticulous, dependable, and persistent. This is partially because they feel that emotions are threatening and confusing, and their strategy of working with this is to remain disciplined, self-restrained, and

proper. They are typically overly polite and even ingratiating toward authority figures but, in contrast, may be somewhat disdainful toward subordinates.

Treatment Implications

The two major goals when working with persons with Schizoid elevations are to (a) encourage at least some increase in social interaction and (b) help them to enhance their ability to experience pleasure. However, these goals are difficult to achieve in a client who is neither likely to become particularly involved in the therapeutic relationship nor ready to place much value in exploration and insight. As a result, the prognosis is poor. In addition, many therapists are likely to feel that schizoids are not particularly rewarding to work with. Therapists must be prepared for long silences and a distant relationship. Yet, any relationship that does develop can be extremely important for the client. Problem solving should be directed at concrete, practical matters. Useful techniques might be audiotape or videotape feedback of their behavior, and cognitive monitoring and reorientation of their internal processes. On the other hand, operant conditioning might prove difficult because they have little capacity for external rewards. Similarly, insight might be unproductive because they are not particularly psychologically minded.

Avoidant (Scale 2A)

Both schizoids and avoidants live solitary, often isolated lives. However, schizoids are indifferent to relationships whereas avoidants desperately want to become accepted and involved with other people—a desire that is blocked by an intense fear of being rejected and humiliated. They warily scan their environment for threats and continually try to present themselves in as favorable a manner as possible. This is rarely successful in that they feel a continual sense of unease, disquiet, anxiety, and overreaction to minor events. Thus, they are frequently preoccupied with intrusive, fearful, and disruptive thoughts. They perceive themselves as socially inept, inferior, and inadequate, and they continually undervalue their achievements. In addition to fear and self-criticism, they frequently feel alone, empty, and isolated. To protect themselves from these fears, they restrict their social environments, constantly maintaining their distance and privacy. This is unfortunate because it undercuts future opportunities of enhancing relationships and places them in a solitary world where they are more likely to reactivate memories of past social rejections. In addition, they rely extensively on fantasy gratification of their needs for affection and anger. Given these dynamics, they are quite likely to fulfill the formal criteria for a social phobia and are frequently depressed. They are frequently described as withdrawn, insecure, edgy, fretful, insecure, isolated, and rejected.

The positive side of avoidants is that they can be extremely sensitive to the needs and perspectives of others. They can potentially show considerable compassion and understanding, and can be emotionally responsive.

Frequent Code Types

Avoidants, along with borderlines, are likely to experience a wide variety of Axis I-related disorders. As a result, it is quite common to see elevations in several of the clinical syndrome scales. Among the most frequent associated disorders are generalized anxiety, phobias, and social phobias (check Anxiety). Depression (check Dysthymia and

Major Depression), hypochondriacal syndromes, and conversion disorders can also occur (check Somatoform). Personality Pattern scales that can be elevated are Dependent, Schizoid, Passive-Aggressive (Negativistic), Narcissistic, and Antisocial. A corresponding elevation on Dependent augments the core dynamics of the avoidant in that the person has even stronger needs to not only become involved with others, but also be supported by and given guidance by them. Avoidant, in combination with Schizoid, adds the dimension of having a lack of awareness or even of interest in personal feelings. These persons are also likely to be detached, aloof, and apathetic, and they rarely develop strong emotional ties with others. They might have some acquaintances, but they are not likely to have any intimate friendships. Elevations on Passive-Aggressive (Negativistic) suggest moodiness and resentment combined with significant difficulty in trusting others. They might vacillate between being friendly and cooperative and then being hostile, which might be followed by apologies. Because they would feel uncomfortable with their anger, they might resort to covert expressions of hostility, such as passive obstructionism. Whereas many persons with avoidant characteristics have low self-esteem, avoidants with elevations on Narcissistic have an inflated sense of importance, overestimating their own value. They are unappreciative of others and justify this attitude by perceiving themselves as special. Situations are framed in such a way as to enhance their own self-worth, and they describe themselves as intelligent, sophisticated, outgoing, and charming. However, their underlying style is avoidant, so their sense of self-importance is extremely flimsy and easily deflated. Elevations on Antisocial introduce to the avoidants' personality a competitive edge that might be expressed in hostile and exploitive behaviors. They would justify this by fears that others are trying to take advantage of them. They usually describe themselves as self-reliant, strong, realistic, and assertive, and they exhibit a contemptuous attitude toward persons who do not have these qualities. In addition, they are likely to be impulsive, argumentative, guarded, reserved, intimidating, cold, and insensitive to the feelings of others.

Treatment Implications

Avoidants are among the most frequent clients in therapy. A potentially difficult issue is that they reveal only the information they believe will not lead to rejection by the therapist. The central treatment task is to change these clients' self-image, but this involves working with interpersonal behavior and helping regulate their mood. Particularly useful techniques would be in vivo exposure based on a graded hierarchy, anxiety management training, cognitive reorientation to challenge thinking errors, assertiveness training, and, possibly, psychopharmacological interventions to deal with anxiety states and possible panic attacks. However, the most difficult challenge is to keep them in therapy long enough to achieve therapeutic gain. This would require carefully balancing support, empathy, and trust-building while still encouraging them to experience situations that challenge them to work on new behaviors and perceptions. Because their high level of arousal would be the primary reason for their terminating prematurely, techniques of arousal reduction—emotional support, reassurance, relaxation, hypnosis, thought stopping, and supportive interpretation—would be particularly important to use. Typically, these clients make significant therapeutic gains. One area to investigate is the possibility that they are using alcohol to medicate their anxiety. However, referral to Alcoholics Anonymous might be difficult, given their avoidant style; therefore, other forms of intervention should be considered.

Depressive (Scale 2B)

The depressive personality style involves not merely recurrent symptoms of depression, but an enduring pattern of thoughts, attitudes, behaviors, and self-concepts related to depression. These clients perceive themselves as worthless, vulnerable, inadequate, unsuccessful, and guilty, and they frequently engage in self-criticism. When possible, they frame events in a defeatist, fatalistic manner. They have learned to expect ridicule and derision. Even extremely slight signs of indifference might be interpreted as contempt and condemnation. Others perceive them as forlorn, somber, discouraged, and hopeless. They similarly describe themselves as discouraged, quiet, drained of energy, and despairing. Initially, their depressive behavior might elicit support and empathy from well-intentioned others. Eventually, however, they end up feeling deserted and abandoned because their interpersonal behavior is likely to either distance others or attract persons who will use their passivity and depression to exploit or otherwise control them. They rarely engage in active, assertive behavior to obtain reinforcement from others. They feel powerless and at the control of forces beyond their control. Although they crave love and support, they fail to act in ways that others find attractive and gratifying. Sometimes their self-criticism is a tactic to diffuse the potential criticism of others and simultaneously solicit support and sympathy. As a result, their interpersonal style serves to further reinforce their depression, and they frequently end up feeling angry, resentful, and pessimistic. Depressive personality disorder can be distinguished from major affective disorder and dysthymia in that, with a depressive personality, there will be an early, extended onset (versus more rapid and intense), along with multiple personality traits consistent with depression.

Because depressives are quite introspective, they have the potential for and the orientation toward developing depth of insight. In addition, they are emotionally responsive and often have depth of feeling. Their level of distress may also be used as an aid in motivating them to change. High-functioning depressives may be able to have genuine close, caring relationships with others and may be articulate, conscientious, responsible, and insightful. They might potentially respond well to humor, elicit liking in others, and be able to effectively take into account alternative points of view.

Frequent Code Types

The most likely elevations on the clinical scales would be on Dysthymia, Major Depression, and, possibly, Bipolar: Manic. These would be natural extensions of the individual's overall depressive style. Considerable conceptual and clinical overlap with other personality scales is likely, resulting in frequent associated elevations on Schizoid, Avoidant, Passive-Aggressive (Negativistic), Self-Defeating (Masochistic), and Borderline. An associated elevation on Schizoid would introduce an apathetic, indifferent, self-sufficient element to the depressive style. Because they are more likely to be interested in inanimate objects than interpersonal relationships, developing an effective therapeutic working relationship would be difficult. Organizing and logically communicating thoughts is often extremely difficult. If Depressive and Avoidant are both elevated, the depressive style is characterized by anxiety and fear of interpersonal humiliation, which leads to isolation as they attempt to protect themselves. They engage in extreme introspection and have a sense of alienation from themselves. They are inhibited, have few social skills, feel easily embarrassed, and have few close friendships. They are also

likely to have difficulty experiencing pleasure and feel inhibited about pursuing goals. An elevation on Passive-Aggressive (Negativistic) flavors the depression with anger, irritability, and sour grumbling. They vacillate between being bitter and resentful toward others and being intropunitive and self-deprecatory. Because they are uncomfortable with their anger and resentment, these feelings are typically expressed in indirect ways, such as through obstinacy, procrastination, and inefficiency. There are clear similarities between Depressive and Self-Defeating (Masochistic). Both of these scales emphasize behaviors that result in the person's not obtaining what he or she wants from life. However, elevation in both of these scales highlights active maneuvers that result in possibly undeserved blame and unjust criticism. These persons present themselves as self-effacing, self-sacrificing, obsequious, and deserving of painful consequences. They are likely to get drawn into relationships that are physically or emotionally abusive, but they respond by being ingratiating and submissive. A Borderline and Depressive configuration emphasizes a serious difficulty with controlling affect and behavior. Cyclical variations of emotional constraint and criticism are followed by impulsive outbursts, sometimes of a self-destructive nature. Suicide potential needs to be carefully monitored. They are likely to have difficulty comforting themselves when distressed and feel that life is meaningless. Problems are typically expanded out of proportion (catastrophized). Accusations may be made that others have mistreated them. Their level of self-identity is extremely weak, and sometimes they have difficulty logically organizing their thoughts and emotions.

Treatment Implications

The major focus of intervention should be to work with their sense of helpless immobility and their belief that emotional pain is an inevitable life condition. Interventions related to interpersonal behavior, cognitive schemas, self-concept, and expectations are often essential. Specific techniques might include social skills and assertiveness training, cognitive interventions that challenge underlying assumptions, behavioral programs that enhance pleasure-related activities, and group involvement that combines support and encouragement for change. Initial contact should be characterized by support that seeks to satisfy some of the client's dependency needs without fostering further helplessness. Psychopharmacology might be considered but should not be an end in itself. Long-standing cognitions, modes of interacting with others, and self-concept persist even after medication might have removed some of the more symptomatic features of the disorder. Therapeutic challenges involve preventing self-harm, preventing the client from proceeding too fast and possibly encountering failure and disillusionment, and preventing relapse. Relapse prevention can be enhanced by realistically advising that some recurrent difficulties are inevitable.

Dependent (Scale 3)

The core characteristic for persons with elevations on this profile is their feeling incapable and incompetent of functioning independently and, therefore, unable to create strong bonds with people whom they perceive as being able to lead and care for them. They quickly create alliances and give up responsibility for decisions. Thus, they feel inadequate and insecure, and they have low self-esteem. They usually describe themselves as placating, insecure, passive, immature, and deserted. A primary way in which they

deal with these feelings is to identify with stronger people and define themselves in terms of these people. They are continually concerned with the possibility of losing friends. To maintain friendships, they are extremely submissive and cooperative, and cover up any unpleasant emotions out of fear that the emotions might alienate others. They, therefore, minimize objective problems, rarely disagree with others, and never take a strong position on an issue. Others, therefore, perceive them as gullible, wishy-washy, humble, timid, docile, and passive. Internally, they have a limited range of competencies in reducing tension and stressors. Elevations on this scale are consistent with bulimia (check Noteworth responses related to eating disorders: items 121, 143, 155, 163).

Often, dependent personalities are well-liked because they are cooperative, compliant, and humble, and they value the opinions of others. They are also likely to be loyal, warm, tender, and noncompetitive. They attempt to develop and maintain lasting friendships and do so, in part, by defusing unnecessary conflict.

Frequent Code Types

The most frequent Axis I-related difficulty is likely to be an anxiety disorder (check Anxiety), which might include panic attacks, social phobias, and agoraphobic attacks often related to or triggered by fears of separation. Mood disorders are represented by associated elevations on Dysthymia as well as Bipolar: Manic and Major Depression might also be common. Frequent associated scale elevations on the personality scales include Avoidant, Schizoid (see section on Schizoid), Compulsive, Passive-Aggressive (Negativistic), Histrionic, and Self-Defeating. An associated elevation on Compulsive indicates that dependent characteristics are combined with seeking approval and nurturance from others by acting perfectionistic, disciplined, orderly, industrious, and persistent. They are highly respectful and even ingratiating toward persons in positions of authority. Careful preparation will be made for future events. As a result of their dependency and focus on details, they likely have difficulty making decisions. Elevations on Passive-Aggressive (Negativistic), along with Dependent, indicate that, although these persons seek the guidance and leadership of others, they are also quite conflicted about these relationships. They may vacillate between appearing to cooperate and then feeling resentful and angry, which leads to resistance toward others in power. Guilt follows, but then the cycle is likely to repeat itself. High scores on Histrionic indicate that these clients are active and outgoing in attempting to get others to notice and take care of them. To this end, they might appear charming, dramatic, seductive, and extroverted. They are often quite sensitive to the moods of others, but may have noteworthy difficulty and a feeling of emptiness when they have to act independently. Finally, when Self-Defeating is elevated with Dependent, it highlights these clients' poor self-esteem, based in part on having been in a series of relationships that have been painful. Although they desperately want others to care for them, they present themselves in a negative and pessimistic manner. Eventually, they undermine and sabotage the relationships that, on another level, they seek to create.

Treatment Implications

Dependents frequently seek treatment. Typically, rapport is quite easily established, especially if the therapist responds in an authoritative, comforting, and assertive manner. However, the greatest danger (or challenge) is that a relationship may be created in which the therapist becomes a rescuer, thereby reinforcing the dependent pattern.

These clients may prefer the therapist to be directive, but a nondirective, Socratic method is more likely to encourage assertion and independence. An important goal is to reduce their clinging patterns and, instead, encourage their interacting in a more direct, assertive manner. Specific techniques might include assertiveness training, anxiety reduction skills (deep breathing, muscle relaxation, meditation), role playing, group therapy (to explore their impact on others), and psychoanalytic techniques that can probe the origins of their dependent patterns.

Histrionic (Scale 4)

Histrionic persons are dramatic, colorful, and emotional. Their tolerance for boredom is extremely low, and they are constantly seeking new situations. By focusing on the external world, they do not fully digest and integrate their experiences with their inner world. Because experiences are not integrated, they do not grow and learn from them. As a result, their level of maturation does not progress. They typically become highly invested in situations or with friends, but, when the excitement ends, they reinvest their energy and interest elsewhere. They typically describe themselves as active, egocentric, exhibitionistic, flighty, extroverted, and flirtatious. They also see themselves as charming, outgoing, and able to acquire the attention of other people. As a result, they make very good impressions in party-type situations, although sometimes they might be perceived as too loud, demanding, and uncontrollable. In addition, they might be exhibitionistic and seductive, placing excessive reliance on physical appearance. Because they react easily and spontaneously to new situations, it is easy for them to mingle with people and quickly establish friendships. However, behind these seemingly assertive and independent behaviors are strong needs for dependency. Whereas dependents seek the protection and guidance of others, histrionics also need the attention and support of others but seek it in an extroverted, overt manner rather than using more submissive methods. Behind histrionics' dramatics and high level of activity are often conflicted, painful feelings that they avoid focusing on. Thus, their activity allows them to skim the surface of these feelings. Dissociative techniques, including the development of conversion reactions, may even be used. Typically, they communicate in a global, general manner in which they make arbitrary judgments with little focus on the specifics of an event or concept.

Histrionics can be warm, colorful, interesting, engaging, and emotionally responsive; typically, they have a good sense of humor. They easily adapt to new situations and, at least superficially, appear to have little difficulty interacting with and becoming close with others. Elevations on Histrionic are associated with an above-average number of positive life events, low levels of distress, and good social adjustment. As a result, if Histrionic is the only elevated scale, it should not be used to suggest a personality disorder, but more a style of adapting.

Frequent Code Types

Because of their underlying feelings of dependency, histrionics are likely to experience separation anxieties or, as an expression of their fears of emptiness, agoraphobia (check Anxiety). Conversion symptoms or hypochondriasis might also be a means of dramatically expressing their needs (check Somatoform), and their need for stimulus

seeking may result in substance abuse (check Alcohol Dependence and Drug Dependence). Possible associated elevations on personality scales include Dependent (see section on Dependent), Narcissistic, Passive-Aggressive (Negativistic), Antisocial, and Compulsive. Elevations on Somatoform might indicate conversions. An associated Narcissistic elevation, along with Histrionic, frequently occurs with and is quite consistent with Histrionic in that it exaggerates many of the self-centered qualities of the histrionics. They are also likely to emphasize how charming and capable they are and to belittle those who do not partake in reinforcing their own sense of self-importance. Their descriptions of their competence and exploits are often exaggerated. They continually indicate how they are special and worthy of more attention and praise than others. An associated elevation on Passive-Aggressive (Negativistic) is problematic in that the histrionics do not like to accept their own negative emotions, such as anger and resentment. As a result of this conflict, they are moody, unpredictable, and emotionally reactive. They might overtly criticize or show disdain for others or, in contrast, express these feelings in a more indirect way, such as through obstructionism. Their attempts to repress and overcontrol their anger and resentment may sometimes culminate in explosive outbursts, followed by guilt and apologies. Similarly, an elevation on Antisocial creates conflict for these persons. They are highly dependent on others, but they also realize that their anger, disaffiliation, and resentment are likely to distance the very people whom they so much need. They might begin a relationship by being charming, friendly, and engaging, but eventually their antisocial feelings become expressed in resentment, mistrust, and even anger. In extreme cases, they might fluctuate between overcontrol and occasional extreme emotional or even physical outbursts. They may also seek to cope with this conflict through passive-aggressive strategies. Their world is perceived as a competitive, potentially dangerous place, and, given these perceptions, they have similarly become competitive, tough realists who believe that this is the only means of coping. Elevations on Compulsive, along with Histrionic, also present a conflicted relationship because part of the person wants to be unrestrained and emotional whereas another part believes in the importance of emotional overcontrol. These clients are likely to seek approval through being orderly, efficient, dependable, and by dressing correctly. Often, they have difficulty integrating these two modes of adapting and may become tense and moody.

Treatment Implications

Histrionics are typically motivated to come to therapy because they have been through a time when they have been criticized and feel socially deprived. They describe feeling empty, bored, lonely, and discontented. Because they are emotional, responsive, friendly, and seek the support and approval of others, they are likely to become easily engaged in therapy. These qualities usually lead to an initial high level of motivation and a good prognosis. They are unlikely to develop severe or chronic forms of psychopathology. However, they usually stay in therapy only long enough to become stabilized and rarely engage in deeper levels of self-exploration. One of the primary goals is to reduce their overdramatization. A calm, objective, cognitive approach is often useful in achieving this goal. In addition, group or family interventions can be useful in enhancing and practicing improved interpersonal skills. Given their externalizing coping style,

a behavioral approach, combined with the development of specific skills, is likely to be more effective than one attempting to develop extensive insight.

Narcissistic (Scale 5)

The central characteristic of individuals with elevations on this scale is their exaggerated sense of self-importance and competence. Because they perceive themselves as special, they are likely to assume that many of the conventional rules of living with people do not apply to themselves. In addition, they may feel that they deserve special favors without having to reciprocate the time and resources that are given to them. As a result, they are likely to be overrepresented among persons seeking workers' and other forms of legal compensation. Internally, they might be quite creative in developing plausible reasons for their self-centeredness, but, to others, these reasons might seem flimsy and transparent. Their fantasies typically involve immature, self-glorifying situations in which they are the center of attention because they are beloved, admired, successful, and physically attractive. In real life, failures are quickly rationalized and conflicts are minimized, and they are adept at enhancing their sense of pride. In building their image, they might depreciate the value of others to make themselves look superior by comparison. They might, therefore, appear arrogant, haughty, snobbish, pretentious, and conceited. They present themselves as intelligent, sophisticated, outgoing, and charming, with an air of cool optimism and feigned tranquility. Rarely do they express any self-doubt. Interpersonally, they are likely to be exploitive, autocratic, and insensitive to the needs and feelings of others. Thus, they are generally lacking in empathy. They constantly attempt to obtain admiration from others. If they are in situations in which they are criticized, they might become quite competitive and aggressive toward those who criticize them, or they may react with contempt or indifference. Thus, they have a primarily externalizing coping style. If their narcissistic bubble is burst, they are at risk for becoming depressed and potentially involved in substance abuse. A subgroup of high scorers are well adjusted and do not experience much emotional distress (see the following positive descriptions). As such, high scores should be interpreted as merely a style of adapting rather than a possible disorder. In contrast, others are pathologically narcissistic. Thus, a diagnostic challenge is determining in which of the two groups the client best fits. If Narcissistic is only mildly elevated and the only elevation, it strongly suggests that they are likely to be in the well-adjusted group.

They frequently make excellent first impressions and might even receive respect and affection from others. Typically, they are articulate, carry themselves with dignity, and have a good sense of humor. Others often perceive them as being proud, independent, confident, and optimistic.

Frequent Code Types

Because narcissistic persons are prone to affective disorders and substance abuse, check relevant clinical scales (Bipolar: Manic, Dysthymia, Alcohol Dependence, Drug Dependence, Major Depression). Personality scales that are likely to be elevated include Avoidant (see section on Avoidant), Histrionic (see section on Histrionic), Antisocial, and Passive-Aggressive (Negativistic). Elevations on Narcissistic and Antisocial emphasize the self-centered, competitive, and possibly aggressive and intimidating

character of these persons. They are likely to be hostile and exploitive and justify this conduct by pointing out the competitive and exploitive nature of other people. At times, they might become malicious, cruel, and abusive; at others times, they may be cheerful, gracious, and friendly. Because they fear the criticism and possible exploitiveness of others, they might frequently be guarded, resentful, and reserved. The combination of Passive-Aggressive (Negativistic) and Narcissistic places these persons in a difficult, conflicted position. They seek to perceive themselves as superior and special in relation to others, but they are also acutely aware of their limitations. Thus, they are likely to be apologetic, submissive, compliant, and cooperative on the one hand, but also hypersensitive, moody, resentful, and angry on the other. They have marked difficulty in accepting criticism, combined with frequent mood changes.

Treatment Implications

Because attending therapy is an implicit admission of imperfections, it is unusual for narcissistic persons to initiate therapy themselves. When they do, it is usually because their narcissistic sense of superiority has been compromised through events such as divorce or loss of employment. Interpersonally, they are likely to remain aloof and often be competitive with the therapist. They might question how someone who is less talented than they are could possibly be of assistance. Alternatively, they might elevate and inflate the status of the therapist because their association with someone who is so accomplished can be used to bolster their own sense of self-esteem. The easiest tactic for returning them to their previous level of functioning is to encourage and support them in recounting their previous successes and achievements. However, this may do them a disservice in the end because they will not learn new strategies of coping and relating. A particularly useful technique might be cognitive reorientation, in which they are helped to challenge the need to be perfect and desensitized to criticism. Group and family therapy might support them in achieving more realistic and adaptive interpersonal skills. Given that they are likely to deny imperfections and resist change, either paradoxical interventions or approaches that use nondirective or self-directed techniques are likely to produce the best outcomes.

Antisocial (Scale 6A)

The central theme for persons with elevations on this scale is competitiveness along with impulsive acting-out of antisocial feelings. They are often described as provocative, violent, vicious, self-centered, dominant, dishonest, brutal, and devious. Their actions are often hasty, short-sighted, imprudent, and they generally ignore the consequences of their actions even to the extent of disregarding the safety of themselves and others. They can be interpersonally irresponsible—they will violate the personal rights of others in occupational, marital, parental, or financial contexts. They can be expected to have legal difficulties because many persons with elevations on Antisocial engage in criminal activities. For others in this category, legal problems are often absent because they confine their acting out to legal domains such as alcohol abuse, interpersonal insensitivity, unreliable work practices, and irresponsible sexual behavior. However, they do not conform to social norms and may even feel and express contempt toward these norms. They enjoy the feeling of not being confined by standard modes of

conduct and project the image of being free, flexible, unencumbered, and having little obligation to schedules, commitments, or persons. Unfortunately, this image is usually associated with a lack of compassion, empathy, remorse, and charitableness. Frequent expressions of callous competitiveness are justified by their pointing out the exploitiveness of others, or otherwise conceptualizing the world as functioning according to the "law of the jungle." Because of these attitudes, they are mistrustful, suspicious, guarded, and reserved. They might also be aggressive, intimidating, cold, insensitive, or even cruel and malicious, thereby provoking fear. Those who are considered "weak" may be treated with contempt, or their own malicious tendencies might be ascribed to others. When challenged, they are likely to become impulsively angry or resentful, vindictive, and vengeful.

At their best, antisocials can be gracious, charming, friendly, and cheerful. Some people might perceive them as interesting and exciting, at least in part, because they are not confined by the same rules of conduct and restraints that other people have.

Frequent Code Types

Check to see whether the clinical scales of Alcohol Dependence and Drug Dependence are elevated; given the impulsiveness and hedonism of antisocials, they are prone to substance abuse. Although generally free from anxiety, they can develop affective disorders, especially when being held accountable for antisocial acting-out (check Bipolar: Manic, Dysthymia, Major Depression). Associated personality scales that are frequently elevated include Avoidant (see section on Avoidant), Dependent, Narcissistic (see section on Narcissistic), Histrionic (see section on Histrionic), Compulsive, Passive-Aggressive (Negativistic), and Aggressive (Sadistic). High points on Antisocial and Dependent indicate that these persons are extremely conflicted because they perceive the world as a difficult, competitive place; yet at the same time, they feel that they need to rely on others for protection and guidance. They are mistrustful, guarded, and reserved, and, although they know that to function they need to be tough, they do not feel themselves capable of this stance. The combination of Antisocial and Compulsive is also a conflicted combination. These persons feel internally impulsive, but they believe in discipline, control, persistence, and dependability. Their typical strategy is to become emotionally overcontrolled, careful, and deliberate. Perceiving the world as competitive and potentially exploitive, they protect themselves with a strategy of hard work, self-restraint, thorough preparation, and being guarded and mistrustful. Others are likely to perceive them as emotionally distant, tough-minded, formal, perfectionistic, inflexible, and possibly indecisive. When Passive-Aggressive (Negativistic) is high along with Antisocial, the angry, resentful characteristics of the antisocial are brought out; yet, the same individuals may desire the closeness and warmth that could be available in relationships. However, they perceive the world as a struggle in which most situations are framed in "win–lose" terms. Thus, they frequently override their need for affection by becoming tough-minded, competitive, and interpersonally superficial. They might excel in individualistic activities—some competitive sports or sales positions, for example—but they would have difficulty working in situations that require loyalty and team coordination. The unusual combination of Antisocial and Aggressive (Sadistic) is noteworthy as it indicates that any acting-out will be cruel, malicious, and callous. The elevation on Aggressive (Sadistic) indicates that the expression of antisocial feelings is direct, overt, and abusive. Such persons should be treated with considerable caution.

Treatment Considerations

Antisocials typically do not perceive the need for treatment and are most frequently referred either by the courts or because of threats from spouses that they will leave them. Once in therapy, they are likely to either openly defy therapist interventions or develop a facade of cooperation in the hope that they might be able to somehow exploit the situation. Therapists need to be cautious; they can potentially be conned by these clients, who would then perceive them as weak and not worthy of respect. The therapists may then run the risk of becoming angry, cynical, and punitive—and ineffective. Given that the antisocial's style is one of externally acting out, the most appropriate interventions are ones that are directed toward changing specific forms of behavior with clear limits: behavior modification, behavioral contracting, and external monitoring of behavior. Antisocials are unlikely to be responsive to internalizing, insight-oriented interventions. In addition, because their arousal level is typically low, techniques that increase arousal, distress, or even anxiety serve to increase their level of motivation. A group context might work particularly well, because antisocials are more responsive to peer influence than to authority-directed influence. However, most interventions have not been demonstrated to be effective in changing their underlying personality structure. A more realistic goal is the reduction of specific targeted symptoms or behaviors, particularly their aggression, destruction, impulsiveness, and poor affect. Target behaviors might be framed in the context that change is in the client's self-interest.

Aggressive (Sadistic; Scale 6B)

Individuals scoring high on Aggressive are typically competitive, energetic, hardheaded, authoritarian, and socially intolerant. They are predisposed toward aggressive outbursts, which might be expressed in a callous manner with little awareness of the impact of their verbally or physically aggressive actions. In many ways, this can be seen as a further pathological variation of the antisocial personality. Being in control and exerting power perhaps to the point of intimidating others is a central means they use to achieve their goals. Humiliating their victims also serves to release their own psychological pain. Sometimes, they enter socially approved enforcing roles in which their expression of aggression is disguised behind socially sanctioned rules (the strict disciplinarian school principal or overzealous police officer). They are relatively unaffected by pain and punishment and may act in a manner that is both reckless and daring. They have a tough-minded orientation, which might be expressed in a caustic and contemptuous attitude toward social events, and is consistent with their prejudice, intolerance, and authoritarianism. At their worst, they might express vicious, explosive, violent, and even brutal behavior. Noticeably absent is a sense of shame, guilt, sentimentality, or internal conflicts. Other persons are perceived as objects to manipulate and control. This attitude might be enhanced and justified if the victims can be considered members of disempowered, marginalized groups.

A positive aspect of persons with this profile is that they can effectively cope with challenges. They can be unflinching and daring, which, if expressed in the right context, can be considered courageous. In reaching a goal, they are relatively unencumbered by subtle ambiguities that might make it difficult for other people to take action.

Frequent Code Types

Fortunately, elevations on Aggressive (Sadistic) are infrequent but, when they do occur, noteworthy elevations on other scales include Antisocial (see section on Antisocial), Narcissistic, Compulsive, and Paranoid. When Aggressive (Sadistic) is added to Narcissistic, these individuals do not only have an inflated, unrealistic sense of themselves, but also they are likely to be openly hostile and destructive, which is not the case when Narcissistic is elevated by itself. Elevations on the Compulsive scale highlight a methodical and disciplined expression of aggression. A corresponding elevation on Paranoid indicates that these persons' cruelty might be self-justified by suspicions that others would like to exploit or even brutalize them.

Treatment Implications

This difficult-to-treat group almost never reports to therapy on their own initiative. Once in therapy, they are likely to belittle the therapist and may even be overtly hostile. A therapist who responds negatively is likely to be perceived as weak, and they use this perception to discount therapist interventions. In addition, they typically lack insight into their behavior and can even be indifferent to the damage they inflict. Cognitive interventions are unlikely to be successful because their thought patterns are quite rigid. Potentially useful approaches might be anger and impulse management programs, developing assertive as opposed to hostile communications, and persuading them to see that changing some of their more problematic behavior is actually in their own self-interest.

Compulsive (Scale 7)

The core characteristics for persons with this elevation are conformity, discipline, self-restraint, and formality. They strictly adhere to social norms and may even be upset by novel ideas, especially if they challenge established norms of conduct. They are conscientious, well prepared, righteous, and meticulous; and they perform well when required to work on a schedule. They typically work hard, sometimes to the exclusion of leisure activities. Their emotions and behavior are tightly controlled. Interpersonally, they are formal, moral, perfectionistic, and rigid. They are overrespectful and even ingratiating toward persons in authority. In contrast, they are likely to be demanding, perfectionistic, and even contemptuous of subordinates, insisting that they act in strict adherence to correct and preestablished rules and methods. Self-descriptions include responsible, dependable, orderly, punctual, reliable, and stubborn. Internally, they are rigidly controlled and do not allow themselves to experience any forbidden thoughts or impulses. Their world is constructed in terms of schedules, deadlines, rules, ethics, and prescribed forms of behavior. Although they perform well in structured, concrete working environments, they have difficulty adjusting to changing work situations that require creative, spontaneous responses. Although these strategies provide them with a high degree of control over their world and their inner impulses, the price they pay is a grim, tense, joyless life in which warm feelings and spontaneity are kept under tight control.

Positive qualities include loyalty, prudence, consistency, predictability, and a strong sense of duty. Often, they are able to approach a difficult situation with maturity and competence. In a work context, they are punctual, thorough, diligent, honest, and rarely

make mistakes. Often persons with elevations on Compulsive are high achievers and rarely report psychiatric distress. The evidence is very strong that this scale does not measure aspects of disorder but, rather, it reflects a person's adaptive style. For example, it has been found that persons who have been formally diagnosed with an obsessive-compulsive disorder rarely had significant elevations on the Compulsive scale.

A defensive, fake good profile can produce an elevation on Compulsive. In these cases, the previous scale interpretation should not focus on discipline and restraint but rather on the client's defensiveness.

Frequent Code Types

Typically, elevations on Compulsive are not accompanied by elevations on other scales. However, when comorbid conditions do occur, the most frequent Axis I problems are generalized anxiety disorders (check Anxiety scale) and depression, particularly of an agitated nature (check Dysthymia and, possibly, Major Depression scales). Compared to other personality disorders, compulsives tend to be a better defined population as there is less overlap with other personality disorders. Nonetheless, associated elevations can occur with Schizoid (see section on Schizoid), Dependent (see section on Dependent), Histrionic (see section on Histrionic), Antisocial (see section on Antisocial), Avoidant, and Narcissistic. Concurrent elevations on Avoidant indicate that these individuals would like to obtain the warmth and affection of others. However, they are extremely hesitant to do so because people are perceived as unpredictable and emotional. Both these aspects of relationships are experienced as risky and are likely to arouse significant anxiety. Compulsives have learned to minimize risk by becoming perfectionistic and relating in a distant, aloof manner. Elevations with Compulsive and Narcissistic suggest individuals who are confident, defensive, and unlikely to concede that they have made a mistake. They strongly rely on their own ideas and are likely to have difficulty accepting the advice, suggestions, and especially the orders of others. Individuals perceive them as inflexible, formal, proper, and distant. As a result, they have difficulty working in supportive team environments where mutual respect and consensus building are crucial factors.

Treatment Implications

Usually, compulsives lead controlled, predictable, and generally functional lives. However, when confronted with excessive change or important decisions, they may present to therapy with anxiety-related problems. In particular, these might be expressed in somatic complaints because they have a difficult time releasing internal tension. They often view their world in a rigid, inflexible manner. As a result, self-exploration is difficult because it is experienced as a violation of their "character armor" and their personal sense of privacy and conformity. In addition, self-exploration runs the risk of playing into their obsessiveness, so that change never actually occurs. One technique of breaking up their obsessive patterns is to help them access and experience their affect. Other strategies are to work with them to realize the irrationality of their patterns or to use paradoxical interventions (i.e., reframing perfection as actually allowing themselves to make mistakes). Usually, the first line of intervention is support, combined with techniques of anxiety reduction: systematic desensitization, relaxation, emotional support, biofeedback, and, possibly, psychopharmacological agents. Any insight-related work

should proceed cautiously and with considerable reassurance, so that their defenses are not challenged too quickly. Potentially problematic client–therapist transactions might be therapist boredom, power struggles, or therapist collusion with the client's compulsions in the form of endless but unproductive insights. Despite these potential difficulties, their prognosis for treatment is quite good.

Passive-Aggressive (Negativistic; Scale 8A)

The core characteristic for clients with elevations in this scale is a mix of passive compliance combined with resentment and opposition. These clients usually act on these resentments in impulsive and erratic ways. Feeding their resentment is a sense that they have somehow gotten a raw deal in life and will inevitably be disappointed in relationships. However, they also feel that their resentment and anger are not acceptable emotions for them to have. As a result, guilt and conflict pervade their lives. This internal conflictual style also becomes externalized and creates problems in interpersonal relationships. They are moody, complaining, and intermittently hostile. One moment they might be angry and stubborn, but the next moment they feel guilty and apologetic. They are likely to express their negativism in indirect ways—procrastination, inefficiency, and contrary behavior that has the effect of undermining the happiness of others. They may also act on their resentment with caustic comments, complaints, and expressions of contempt toward others. One means of coping with these feelings is to deny them and, instead, attribute them to others. Another way is to conceptualize that the resentment and anger are justified because of the numerous reasons to be envious toward others, who are constantly seen as having things so much better. Their resulting chronic unhappiness is expressed through pessimism, disillusionment, and cynicism. Because they blame other people for their misfortunes, they have little insight into how their own behavior and attitudes cause others to reject them. However, when their attitudes and behaviors eventually lead to rejection by others, these clients feel demeaned, abandoned, unappreciated, and disillusioned. Thus, their difficulties are self-fulfilling and self-maintaining. They typically describe themselves as moody, testy, resentful, oppositional, and discontented.

A further core conflict is a feeling that they would like to depend on others, but this dependence is neither socially acceptable nor safe because others inevitably exploit and disappoint them. Thus, they seem moody and unpredictable as they ruminate over these contradictory feelings. They often perceive relationships as a threat to their safety. To protect themselves, they become superficially quite self-sufficient and independent.

At their best, persons with this elevation can be agreeable and friendly. They can also be flexible, changeable, emotionally responsive, and sensitive.

Frequent Code Types

Persons with this code type experience frequent rejection and are likely to experience depression (check Dysthymia and Major Affective Disorder). Their feeling that interpersonal situations are potentially dangerous is capable of producing chronic anxiety (check Anxiety), which might be expressed in indirect ways through psychophysiological disorders or conversions (check Somatoform). Concurrent elevations on personality scales include Schizoid, Avoidant, Dependent, Histrionic, Antisocial, Narcissistic,

and Aggressive (Sadistic; check frequent code type descriptions for each one of these scales in previous sections).

Treatment Implications

The two major areas of intervention involve enabling Passive-Aggressives to be more consistent in their approach to life and to develop insight into the nature of their ambivalent style of responding. However, the therapeutic relationship itself is likely to be complicated by their ambivalence. Specifically, they desire caring and support by others but perceive the development of such a relationship as a threat to their independence and fear that it will end up with rejection and disappointment. As a result, they may erratically criticize their therapist or engage in passive resistance. Dealing with this potential difficulty through early behavioral contracting might be particularly useful in keeping these clients engaged in the therapy process. One concern related to clinical management is that their impulsiveness might involve suicide risk. This is especially problematic if they decompensate into an anxiety or depressive disorder. Family and marital interventions are likely to be extremely beneficial because passive-aggressive (negativistic) patterns are both initiated by and maintained in these systems. Formal programs of anger management and assertiveness training might also be quite helpful in developing greater control over impulses and learning more effective styles of communication. Their belief in future disappointments, along with their dysfunctional thoughts of having been cheated by life, can be worked on through cognitive interventions that challenge these assumptions. Because they are likely to be resistant, controlling clients, the use of either paradoxical directives or a combination of nondirective and client-directed techniques is likely to optimize outcome.

Self-Defeating (Scale 8B)

High elevations on Self-Defeating indicate aggrieved persons who continually place themselves in situations in which they will be the victims. They present themselves as inferior, nonindulgent, self-effacing, insecure, or otherwise reluctant to accept pleasure and happiness. Somehow, pleasure is seen as something they do not deserve, and they feel that if they allow themselves to experience pleasure, further difficulties or other unpleasant consequences will follow. Anything positive is expressed with very little enthusiasm. Interpersonal relationships are characterized by these clients as servile, self-effacing, self-sacrificing, or otherwise allowing or even encouraging others to exploit or mistreat them. This active involvement in creating situations in which they will be exploited differentiates these types of persons from other depressed clients. Close relationships are usually associated with disappointments and frustrations. Those who do try to support and help them are likely to be ignored or otherwise rendered ineffectual. One purpose of this response is to make themselves weak and harmless in an effort to discourage possible criticism and aggression from others and evoke guilt instead. In addition, their public displays of dejection initially produce both sympathy and a tacit permission to avoid unpleasant responsibilities. A further purpose is to keep their self-identity organized around being shamed, humbled, and debased. They may be so absorbed in their own suffering and misery that they have few resources for appreciating the dilemmas others might be in. Although, superficially, they might be sympathetic to

others, underneath they are unempathic and distrustful. They focus and ruminate on past failed relationships and disparage any personal achievements. This results in their being anxious, apprehensive, mournful, anguished, and tormented.

Positive qualities are that, in comparison to disorders such as schizoid, they are involved with and connected to people. Often, they can develop a good level of insight into their difficulties. In addition, their level of distress is likely to be sufficiently high that they can and do become engaged in therapy.

Frequent Code Types

The greatest risk for self-defeating persons is the development of depression (check Dysthymia and Major Affective Disorder). If anxiety is present, it is usually diffuse and associated with fears of loss and abandonment. Hypochondriacal strategies might be grafted on to their aggrieved style as a means of channeling anxiety and obtaining support (check Somatoform). The most frequent associated elevations are with Dependent (see section on Dependent), Borderline (see section on Borderline), Depressive (see section on Depressive), and Avoidant. When Self-Defeating and Avoidant are both elevated, it suggests that these persons have found relationships sufficiently painful that they have withdrawn to the extent of rarely interacting and becoming relatively isolated. They would like to be involved with others, but that experience has simply proven to be too painful in the past.

Treatment Implications

The paradox of working with self-defeating persons is that the context of therapy is to make them happier; yet, on one level, they do not want to be happier. These clients might even try to provoke or at least frame situations in such a way that they feel rejected or humiliated by the therapist. To counter this tactic, a sufficient amount of support, understanding, and rapport must be established to work with these clients and make them understand that they do not necessarily have to suffer. Specific self-defeating behaviors need to be identified along with the circumstances that elicit them. Assertiveness training, to help clarify their rights and develop skills to stop exploitation, might be particularly helpful. These skills, and others, might be practiced in the context of role plays and/or couples therapy. Further examination of relationships and the part they play in them can occur both in individual therapy and through supportive group interaction.

SEVERE PERSONALITY PATHOLOGY

Schizotypal (Scale S)

The major characteristics of persons with elevations on Schizotypal are eccentricity, disorganization, and social isolation. These difficulties are usually of a long-term nature. Their eccentricities relate to peculiar mannerisms, strange clothes, and bizarre expressions. They typically look drab, lifeless, apathetic, and joyless. Self-descriptions include alienated, isolated, fragmented, and detached. They may engage in magical behavior and rituals in an attempt to neutralize "evil" thoughts, deeds, or omens. Often,

there is little distinction between fantasy and reality. Their communication style is characterized by tangential comments, personal irrelevancies, and magical associations. As a result, they lead empty and meaningless lives in which they drift to and from various locations and sources of employment. Thus, they exist on the fringe of society. Some are detached and emotionally bland; others are more suspicious, anxious, and apprehensive. Because they are mistrustful and communicate poorly, their relationships usually make them quite uncomfortable. As a result, they develop few, if any, close friendships and prefer privacy and isolation. Usually, they lack the interest and energy to initiate social interaction. Internally, they have a deep sense of emptiness and meaninglessness, which sometimes is sufficiently severe to prompt a full schizophrenic episode. Their thought processes are scattered, autistic, and disorganized. They are likely to have experiences of depersonalization and dissociation. In summary, schizotypals are cognitively impaired in their ability to comprehend interpersonal motivations and communications.

Frequent Code Types

Diagnostically, schizotypals exist somewhere between the less severe schizoid disorder and the more severe schizophrenic disorders. However, there is conceptual and clinical overlap with both these disorders; therefore, elevations on scales that measure these dimensions should be noted (check Schizoid, Thought Disorder, and Delusional Disorder). Accordingly, schizoid and schizophrenic disorders might coexist with schizotypal. The most likely associated elevations on personality scales are Schizoid, Avoidant, and Paranoid. The Schizoid and Avoidant elevations are important in distinguishing two subtypes of schizotypals. An elevation on Schizotypal in combination with Schizoid indicates a more passive, apathetic, detached expression of schizotypal characteristics. These persons are deficient in their capacity to experience emotions, and extremely detached and indifferent toward others. In contrast, an associated elevation on Avoidant indicates a desire for personal contact, but these individuals are more anxious and apprehensive, and actively protect themselves by disengaging from others. If Paranoid is elevated along with Schizotypal, it highlights these clients' suspiciousness along with corresponding ideas of reference. Although their thoughts might be more organized because of the coherence provided by the paranoid content, they still have the tangential thinking and eccentric behavior that are characteristic of persons with elevations on Schizotypal.

Treatment Implications

The prognosis for schizotypal is not good because of the ingrained, long-standing nature of their patterns and the difficulty of engaging them in the therapeutic process. Treatment goals should be tempered accordingly, with a focus on preventing further social isolation and deterioration. Changing these individuals' environment to encourage an increase in supportive interpersonal interaction might be particularly helpful. A further intervention might be to help them express and clarify their thoughts while simultaneously providing emotional support. Psychopharmacological agents might be useful both in helping to organize their thoughts and in reducing the likelihood of their acting on irrational impulses.

Borderline (Scale C)

The core features of individuals with elevations on this scale are instability and unpredictability of mood and behavior. One moment they might feel dejected and disillusioned; sometime later, feelings of euphoria are followed by a phase of intense anger, irritability, and self-destructiveness—possibly even involving self-mutilation. Their self-destructiveness reflects a severely punishing conscience. In addition, much of their unstable behavior seems to be directed by internal factors rather than a reaction to environmental events. They have marked mood swings, intermittent periods of depression, generalized anxiety, and intense emotional attacks on others, followed by apathy and dejection. Although these behaviors often create significant interpersonal difficulties, these clients are also extremely concerned with maintaining the care and emotional support of others. Although they often elicit rejection, they strongly react to fears of abandonment. They might intermittently idealize people, but their ambivalence eventually gives way to devaluing and criticizing the same people they have previously idealized. Thus, their relationships are characterized by ambivalence, instability, and intensity. Underlying many of these behaviors is an extremely poorly developed sense of identity, which is at the core of their dissolution of controls. Their poorly defined sense of self might eventually give way to feelings of emptiness and to disorganized thoughts. Under stress, they might have transient psychotic episodes. However, these episodes are rarely sufficient to be considered a formal thought disorder, and these clients typically return fairly quickly to their previous levels of functioning. They typically describe themselves as depressed, impatient, tense, irritable, disturbed, and anxious.

Frequent Code Types

The symptomatology of borderlines can be extremely diverse; elevations may appear on any of the clinical scales. However, mood disorders (check Bipolar: Manic, Dysthymia, Major Depression) and substance abuse (check Alcohol Dependence and Drug Dependence) are among the most common complications. In many ways, borderlines can be conceptualized as exaggerations or extensions of the less dysfunctional personality disorders of self-defeating (masochistic), passive-aggressive (negativistic), dependent, histrionic, and/or narcissistic. As a result, elevations on one or more of the scales representing these constructs would be expected and would provide further information on these individuals' underlying dynamics and particular mode of expression. Because such a broad spectrum of behaviors is encompassed by the borderlines, this can be crucial information to attend to. One of the most frequent associated scale elevations is when Borderline is combined with Passive-Aggressive (Negativistic), which emphasizes the conflicted aspect of the borderlines. These clients feel intense dependency, yet are anxious and extremely ambivalent about it. They also feel intense resentment and anger but simultaneously believe that such feelings are unacceptable. These intense polarities might naturally give way to both a disintegration of the sense of self and clearly unstable, unpredictable behavior. Another important combination is Borderline and Self-Defeating, which would highlight these clients' impulsive and self-destructive characteristics. Behind their unstable emotions and behavior would lie a strong underlying sense that they were not worthy of happiness, but instead, they should be exploited and humiliated. Thus, the presence of depression and suicide would be an essential aspect of

case management. Elevations on Dependent and Borderline emphasize these clients' low self-esteem, passivity, and apathy, combined with their need for someone else who will care for them and make decisions for them. A corresponding elevation on Histrionic would underscore these persons' dependency but, instead of being apathetic and passive, they would be outgoing, friendly, manipulative, and emotional. When their defenses are challenged, they might increase their activity and attention-seeking to intense levels, but if this strategy does not work, they may deteriorate into futility and self-destructiveness. When Narcissistic is elevated along with Borderline, it suggests that these individuals' self-inflated sense of importance has collapsed into feelings of shame, insecurity, emptiness, and self-condemnation.

Treatment Implications

Although borderlines are notoriously difficult to work with, they are also more amenable to change than many other personality-disordered individuals. The central, initial goal is to build sufficient rapport so that work can begin on stabilizing their erratic behavior and affect. This might involve a reality-oriented approach emphasizing aspects such as limit setting, sympathy, reassurance, advice, and insight regarding internal processes. Borderlines are capable of such a wide range of dysfunctional behaviors that knowing which one to address can sometimes be confusing. In addition, they are an unusually heterogeneous group. For example, depression, anxiety, depersonalization, disorganized thoughts, fears of abandonment, self-destructiveness, and/or ambivalence may all become areas requiring attention. More than for most other client groups, building a strong therapeutic alliance is crucial in helping borderlines to adjust and cope with their many conflicted forms of acting and feeling. Because many borderlines resist authority-directed interventions, group therapy might be indicated because they are more likely to be responsive to peer influence.

Paranoid (Scale P)

The central issue for persons with elevations on Paranoid is suspiciousness and defensiveness, combined with a feeling of superiority. They are constantly vigilant because they feel others will criticize or deceive them. Innocuous events are perceived as insults or as the workings of a world in which others are trying to control or harm them. They distort their world by interpreting events to fit their idiosyncratic views. Because they feel in frequent danger, they are abrasive, touchy, hostile, and irritable. They are likely to feel bitter toward people who have been successful and to believe that their success has been achieved through dishonesty and possibly illegal activities. This process involves denying their own shortcomings and attributing them to others. Although quick to notice and expand on minor faults in others, they are ignorant of these same faults in themselves. These dynamics are used as a means of establishing their own superiority in relation to others. They often describe themselves as misunderstood, righteous, suspicious, mistreated, and defensive.

If high scorers on Paranoid perceive that anyone is trying to control or influence them, they consider this a personal encroachment on their independence and will attack and humiliate the encroacher. As a result, they frequently induce fear and exasperation in others. Unfortunately, their system of making sense of the world is self-fulfilling.

People react negatively to their being mistrustful and even hostile, which provides evidence that indeed the world is a dangerous, insecure place. This pushes them progressively into a more insular world in which their thinking is extremely rigid. The rigidity and insularity are maintained because they depend on their own internal processes for both stimulation and reinforcement. They are terrified of being dominated and consider any sign of dependence an indication of weakness and inferiority. They insist on being the designers of their own fate and, to do so, need to be free from entanglements and obligations. Behind this separateness is a fear of losing their personal control and sense of autonomy. Thus, their extremely tightly organized and coherent personality and cognitive structure makes them feel emotionally and physically disconnected from others. In more extreme cases, these persons may have delusions of grandeur, ideas of reference, and intense fears of persecutory plots.

Frequent Code Types

Given the mistrust and fear expressed by many paranoids, anxiety is probably the most frequent Axis I complication (check Anxiety). Additional difficulties are likely to be obsessive-compulsive syndromes in which they engage in compulsive activities in an attempt to make their world "safe." In severe paranoid states, psychotic symptoms, expressed through delusions and hallucinations, may be present (check Delusional Disorder and Thought Disorder). Related elevations on personality scales include Narcissistic, Passive-Aggressive (Negativistic), Sadistic, and Avoidant. If Narcissistic is elevated, it suggests that, at some earlier stage, these clients' self-inflated sense of importance and superiority has been severely challenged. Paranoid processes become a means to resurrect these beliefs in a way that is further separated from reality and, therefore, requires more drastic measures. The result might be extravagant plans to defend the world from evil, create new societies, or solve insurmountable scientific problems. When an elevation in Passive-Aggressive (Negativistic) is associated with Paranoid, it represents an exaggeration of these persons' faultfinding, resentful, and discontented characteristics. These might be expressed as intense feelings of jealously or as claims that they are being cheated and misunderstood. Because their underlying negativism is unacceptable to themselves, they attribute it to the external world, thereby self-creating the world they are so afraid of. An elevation on both Paranoid and Sadistic suggests that these individuals' paranoia will be expressed in an authoritarian, controlling, intimidating, and belligerent manner. They are likely to ruminate over perceived past wrongs and develop callous plots of revenge. Elevations on Avoidant and Paranoid indicate that these clients are handling their fears and suspicions by becoming progressively more insular, reclusive, and isolated. Insularity helps to protect them from fears that others will be able to influence their thought processes. However, they also feel extremely vulnerable and have serious questions related to their self-esteem.

Treatment Implications

Although paranoid personalities have an intact, organized means of processing their world, they develop and maintain this perspective by insulating themselves from the influence of others and developing extremely rigid cognitive structures. Because

therapy tries both to influence clients and to loosen habitual ways of perceiving the world, these people are difficult to work with. As a result, their prognosis is poor. Furthermore, submitting to therapy is an admission of weakness and of giving up self-sufficiency, and both situations are abhorrent to them. A therapist who is too friendly and empathic is likely to be perceived as being deceitful. High empathy by the therapist has even been found to be counterproductive. In contrast, a therapist who is too distant or who challenges these clients' delusions will seem rejecting. Either approach may, therefore, invoke the clients' suspicions. The relationship requires a delicate balance. Trust needs to be slowly built up with gradual but careful encouragement to perceive events from several different perspectives.

CLINICAL SYNDROMES

Anxiety (Scale A)

High scores indicate clients are complaining of tension, difficulty relaxing, indecisiveness, and apprehension. Additional complaints include a highly sensitive startle response, hyperalertness, and fears related to the onset of poorly defined difficulties. Physiological complaints related to overarousal are also common. These might include insomnia, headaches, nausea, cold sweats, upset stomach, palpitations, excessive perspiration, and muscular aches. Anxiety may be either generalized or more focused, as in social situations or specific phobias. Inspection of responses to individual scale items can help to assess the degree of specificity of the anxiety.

Somatoform (Scale H)

Elevations reflect somatic complaints expressed in areas such as generalized pain, fatigue, multiple vague complaints, and/or preoccupation with health-related difficulties. However, these typically represent psychological conflicts that are being expressed through physical means. If the clients have legitimate physical illnesses, they are likely to be unduly preoccupied and possibly exaggerating their difficulties. In other words, their difficulties are overinterpreted to signify a major illness when the illness is actually relatively minor. Often, the complaints are expressed in a dramatic and/or vague manner. An important function of these complaints is to gain sympathy, attention, or medical reassurance. A careful medical history typically reveals a hypochondriacal pattern in which they are overusers of the health care system.

Bipolar: Manic (Scale N)

High scorers are likely to have mood swings that range from elation to depression. When elated, they are restless and distractible, have an exaggerated sense of self-esteem, and are overly optimistic and impulsive. They have a heightened and general sense of enthusiasm, along with unrealistic goals. Interpersonal relationships have a demanding, intrusive, and pressured quality. There is a reduced need for sleep, erratic

mood shifts, and flighty ideas. Extreme elevations indicate a psychotic process characterized by delusions and possibly hallucinations.

Dysthymia (Scale D)

Elevations on Dysthymia reflect sadness, pessimism, hopelessness, apathy, low self-esteem, and guilt. These persons continuously feel socially awkward, introverted, sad, useless, and filled with self-doubt. Discouragement and a preoccupation with their own inadequacy are also present. They have a sense of futility and may easily break into tears. Somatic complications might include insomnia, a poor appetite or habitual overeating, poor concentration, a continuous sense of feeling tired, and a marked loss of interest in pleasurable activities. Although they may have reduced effectiveness in competently undertaking daily activities, they still remain involved in everyday life. Suicidal ideation might be present and should be investigated further. This, and other details related to the nature of the depression, can be further understood by noting the responses to particular items. Unless the Major Depression scale is markedly elevated, it is unlikely that the depression will be sufficiently severe to be considered psychotic.

Alcohol Dependence (Scale B)

Individuals scoring high on Alcohol Dependence are likely to have had a history of problem drinking. They may have tried to unsuccessfully curb or discontinue their drinking. High scorers are also likely to be having social, family, and/or occupational distress. However, the degree to which their drinking is problematic needs to be assessed in relation to other information on their level of functioning.

Drug Dependence (Scale T)

High scorers will have had a recurring history of difficulties with drug abuse. Also present are a number of traits associated with drug-related difficulties: hedonism, impulsiveness, difficulty conforming to mainstream standards of behavior, self-indulgence, exploitiveness, and narcissistic personality characteristics. High scorers are likely to have difficulty organizing daily life activities and experience social, family, legal, and/or occupational distress.

Posttraumatic Distress Disorder (Scale R)

Elevations on this scale suggest that these individuals have experienced an intense life-threatening event that has resulted in extreme fear, helplessness, and arousal. They have reacted by having uncontrolled, intrusive, and recurrent images or emotions related to the event(s): flashbacks, nightmares, or dissociative feelings that reactivate the event(s). Anxiety-related symptoms might include hypervigilance, hyperalertness, overreactive startle reactions, and a compulsive avoidance of circumstances that might be related to the trauma.

SEVERE SYNDROMES

Thought Disorder (Scale SS)

High scores on Thought Disorder suggest these persons have thoughts that are inconsistent, bizarre, fragmented, and disorganized. In addition, their behavior might be regressed, secretive, and incongruous; and they might be confused, withdrawn, and disoriented. Their affect is likely to be blunted, and they may report hallucinations. Possible diagnoses include schizophrenic, schizophreniform, and brief reactive psychosis.

Major Depression (CC)

High scores suggest severe depression, to the extent that these individuals have difficulty with effective daily living. Psychological difficulties include a sense of hopelessness, suicidal ideation, pessimism, ruminating, and fear of the future. Somatic symptoms might include insomnia, poor concentration, psychomotor slowing or agitation, loss of appetite, weight loss, chronic fatigue, early morning awakening, and loss of sexual desire. They are also likely to feel worthless and to experience guilt. Some high scorers might express their symptoms in an irritable, whining manner, whereas others might be shy, passive, seclusive, and introverted.

Delusional Disorder (PP)

Elevations on this scale indicate acutely paranoid states. These individuals are characterized by irrational but interconnected delusions, persecutory thoughts, and grandiosity. They are hyperalert to possible threats. The most frequent mood is hostile suspiciousness, perhaps to the point of belligerence. They feel mistreated, jealous, and betrayed.

RECOMMENDED READING

Choca, J. P., & Van Denburg, E. (1997). *Interpretive guide to the Millon Clinical Multiaxial Inventory* (2nd ed.). Washington, DC: American Psychological Association.

Jankowski, D. (2002). *A beginner guide to the MCMI-III.* Washington, DC: American Psychological Association.

Millon, T., & Davis, R. D. (1996). *Disorders of personality:* DSM-IV *and beyond.* New York: Wiley.

Millon, T., & Davis, R. D. (2000). *Personality disorders in modern life.* New York: Wiley.

Millon, T. (1997). *The Millon inventories: Clinical and personality assessment.* New York: Guilford Press.

Chapter 9

CALIFORNIA PSYCHOLOGICAL INVENTORY

The California Psychological Inventory (CPI) is a self-administered, paper-and-pencil test composed of 434 true-false statements. The test can be administered either to individuals or groups. Although the test has been used to evaluate individuals between the ages of 12 and 70, it was mainly constructed for use with young adults having a minimum of a fourth-grade reading ability. The CPI items request information concerning an individual's typical behavior patterns, usual feelings and opinions, and attitudes relating to social, ethical, and family matters. The results are plotted on 20 scales and 3 vectors (factors) focusing on aspects of interpersonal relationships that are presented in everyday, commonsense descriptions.

The philosophical orientation of the CPI is based on an appreciation of enduring, commonly discussed personality variables that are relevant throughout different cultures. Thus, it uses familiar commonsense terms such as *dominance, tolerance,* and *self-control,* which Gough has referred to as "folk concepts" (Gough, 2000). Accordingly, it has been translated into more than 40 languages. The value of using such common, easy-to-understand constructs is that they already have "functional validity." In other words, they have immediate cross-cultural relevance, are readily understood by a wide range of people, and have a high degree of power in predicting behavior. This is not to imply that untrained persons can competently interpret the CPI, but rather that the test's roots and original constructs are based on conceptions of human behavior held by most people in most cultures. It is up to the skilled clinician to go beyond these common constructs and into a more subtle, broad, and integrated description of the person. Thus, the test does not have as its primary goal psychometric elegance, nor is it derived from any specific personality theory. The focus and concern of the CPI involve practical usefulness and the development of descriptions that strive to be relevant, understandable, and accurate in terms of behavioral predictions. Because of these assets, it has become one of the more frequently used tests by professional psychologists (Camara et al., 2000; C. Piotrowski & Zalewski, 1993; Watkins et al., 1995); more than 2,000 research studies have either been performed on the CPI or have used it.

The CPI was originally developed by Harrison Gough and published in its original form in 1957. Although reviews of the test have been mixed, most reviewers describe it in favorable terms. For example, Bolton (1992) concludes his review in the eleventh *Mental Measurements Yearbook* by stating that the "CPI is an excellent normal personality assessment device, more reliable than the manual advertises, with good normative

data and outstanding interpretive information" (p. 139). It has been subjected to more than 50 years of research and continuous improvement. The criticisms that have been directed at the CPI have stimulated extensive efforts toward refinement and improvement, including numerous studies on predictive validity, the development of alternate scales, and expanded normative data. Many of these improvements were incorporated into the 1987 and 1996 revisions. For these reasons, the CPI has become a respected and frequently used device in personality assessment, particularly in the areas of career development, personnel selection, interpersonal maladjustment, and predicting antisocial behavior (Gough, 2000; McAllister, 1996).

HISTORY AND DEVELOPMENT

The CPI was developed as an inventory to assess enduring interpersonal personality characteristics in a normal population. Gough published his original scales in 1948, but the first copyrighted edition of the initial 15 scales appeared in 1951. However, it was not until 1957 that a completed set of 18 scales was published by Consulting Psychologists Press. It was further revised in 1987 and two new scales (Empathy and Independence) were included, bringing the total number of scales to 20. These 20 scales measure areas such as social ascendancy, social image, intellectual stance, and conceptual interests. Three of these are validity scales, which assess test-taking attitudes including "fake bad" (*Wb*), "fake good" (*Gi*), and the extent to which highly popular responses are given (*Cm*). Because of a combination of continuing research and a wish to conform to the 1990 Americans with Disabilities Act, the CPI was again revised in 1996. Although 28 items were deleted (bringing the total to 434), the 20 scales and 3 vectors from the 1987 version were retained.

The 1957 version of the CPI was derived from an original item pool of 3,500 questions. Of the 468 items that were eventually selected, 178 were identical to MMPI items, 35 were very similar, and the remaining 255 were developed specifically for the CPI. The items were selected on the basis of both empirical criterion keying and a rational approach in which questions were generated that, from a conceptual point of view, seemed to assess the characteristics the scales were trying to measure. These questions were then given to a sample group and accepted or rejected based on the extent of inter-item correlation. However, the majority of the scales were not developed through the rational approach but rather through empirical criterion keying. Thus, series of questions, which had initially been developed rationally, were administered to different groups having specific, previously assessed characteristics that the scales were eventually intended to measure independently of these groups. Each group was selected using a number of different criteria. For example, ratings by friends and family of an individual's degree of responsibility were used to select a person for inclusion in the sample group for the development of the scale on responsibility. The Achievement via Independence scale was based on college students' grade point averages; the Socialization scale used delinquents and nondelinquents; and Sociability involved the number of extracurricular activities that a student participated in. Items that were found to discriminate between the criterion group (responsibility, sociability, etc.) and a "normal" population were selected for initial inclusion in the scale.

It is important to emphasize that, similar to the MMPI items, the empirical relationships are more important than the "truth" of the content. For example, if a person in the group rated for responsibility answers "true" to the statement "I have never done anything hazardous just for the thrill of it," it does not matter whether he or she has actually performed hazardous behaviors for the thrill of it. The main consideration from a psychometric point of view is that he or she answers "true" to that question, which then indicates the item can be used to help differentiate responsible from nonresponsible persons.

The final step was to cross-validate the items with other populations to determine the extent to which the variable the scale was attempting to measure could be accurately assessed. Of the 18 original scales, 13 used empirical criterion keying, 4 used the rational approach, and the final one (Communality) cannot be easily categorized, although it primarily used a combination of the two techniques. The two new scales in the 1987 (Form 462) and 1996 (Form 434) revisions (Empathy and Independence) used a criterion-keying approach to elicit and score items that already existed in the CPI.

Like the MMPI, the CPI scores are given a standard score (*T* score) with a mean of 50 and a standard deviation of 10. The 1957 scales were standardized on an original normative sample of 6,000 males and 7,000 females having a fairly wide range in age, socioeconomic status, and geographic area. The standardization for the 1996 revision was based on 3,000 participants of each sex selected from the CPI archives to be representative of the U.S. population in age, education, status, and other relevant variables. The 20 scales are arranged so that they relate to the following general domains (Gough, 2000):

- Observable, interpersonal style and orientation (i.e., Sociability, Social Presence).
- Internal normative orientation and values (i.e., Responsibility, Self-Control).
- Aspects of cognitive and intellectual functioning (i.e., Intellectual Efficiency, Achievement via Conformance).
- Measures of role and personal style (i.e., Psychological-Mindedness, Flexibility).

These domains assist interpretation as they help to organize practitioners to provide specific information around these more general domains as well as integrate data with other assessment information. For example, Wechlser IQ scores can be further amplified by noting whether the examinee prefers to work in groups with clear guidelines (conformity) versus more independently. Interpretation is further simplified in that higher scale values are associated with traditionally more favorable qualities and lower scores with more unfavorable qualities (Wallbrown & Jones, 1992). The exception to this is the final scale, F/M, which measures traditionally feminine and masculine characteristics. Gough added three vectors or structural scales to the 1987 version. Rather than organizing these three scales conceptually, he developed them based on factor analysis to measure extraversion-introversion (externality-internality), norm-favoring versus norm-questioning, and degree of self-realization.

The CPI has been put to numerous uses since its initial development in 1957. Megargee (1972) reported that, when the test was first printed, researchers and practitioners used it for many of the more obvious purposes of a psychological test, such as the prediction of scholastic achievement, graduation from high school or college, and performance in specific areas, such as math and English. Later, its uses became much

more diversified to the extent that work has now been done on managerial effectiveness, air traffic controllers, stock market speculators, the degree of creativity in fields such as architecture and mathematics, contraceptive practices, and performance in psychiatric residency programs. Furthermore, cross-cultural studies on validity have been performed in France, Israel, Italy, Japan, Poland, Switzerland, and Taiwan. In the field of counseling, it has been used to predict response to therapy, to aid in the selection of a college major, predict college grade point average (GPA), and to predict the degree of success in graduate education programs such as medicine, dentistry, nursing, and education (see Gough, 2000; McAllister, 1996; Megargee, 2002; P. Meyer & Davis, 1992; Van Hutton, 1990). Surveys indicate that it is one of the more frequently used objective personality tests (Camara et al., 2000).

Whereas previous versions of the CPI provided the options of hand as well as computer scoring, the 1996 (Form 434) can be scored using only computerized facilities (see Consulting Psychologists Press). A variety of interpretive programs are available as well as additional alternate scales. Should practitioners wish to work with the CPI, they can either have protocols scored by Consulting Psychologists Press and interpret the resulting profile themselves, or have the results both scored and interpreted by Consulting Psychologists Press (or other providers). A 250-item short form of the CPI is currently being developed.

COMPARISON WITH THE MMPI

Because of the similarity in both format and item content, comparisons between the CPI and MMPI are inevitable. Thorndike (1959) has referred to the CPI as "the sane man's MMPI," and there are a number of clear similarities. The 1996 version of the CPI is comprised of more than one third of the MMPI's questions (171 out of 434); a conversion is made from raw to standard scale scores with a mean of 50 and standard deviation of 10; and the final values are charted on a graph with peaks and valleys.

Despite these similarities, it is essential for any clinician using the CPI to also appreciate the significant conceptual and psychometric differences between the two tests. The general intent of the MMPI is to assess a person's intrapsychic processes and emotional distress as these relate to specific psychodiagnostic categories. Each of these categories has a group of internal dynamics surrounding it; such as depression, which also includes apathy, lowered capacity for pleasure, and feelings of hopelessness and helplessness. The primary task of the MMPI is to identify either the presence or absence of these internal dynamics and to place the examinee in either a normal or one or more psychopathological categories. In contrast, the CPI focuses more on a normal population and is highly interpersonal in nature. In fact, there is a marked absence of symptom-oriented questions. Thus, the CPI is concerned with the presence or absence of specific interpersonal skills. In addition, the CPI avoids complex diagnostic nomenclature and emphasizes practical descriptions that are commonly used in most cultures.

From a psychometric perspective, the MMPI was originally developed from a bimodal distribution in which the main focus of the test was to classify a specific client in either a pathological group or a normal one. The contrast groups were not high or low on a specific trait, but, rather, were high in pathology when compared with normals. For

example, a group that was high in hysterical traits was contrasted, not with a group of persons having superior health, but with individuals having only an average number of hysterical traits. In clinical assessment, members of the pathological group are considered to be persons scoring greater than 1.5 standard deviations above the norm ($T = 65$). As a result of this emphasis on differentiating pathological groups from average or normal groups, the interpretation of profiles within "normal" ranges (i.e., $T = 35$ to 60) is uncertain and should be approached with extreme caution. In contrast, the CPI uses a normal distribution within a standardized population. Furthermore, Gough used groups whose behavior was extreme on both high and low dimensions of the characteristic being measured. Thus, normal range scores of less than 1.5 standard deviations from the mean can be interpreted with a fairly high level of confidence. For example, a CPI score on *Ac* (Achievement via Conformance) of $T = 60$ indicates a fairly high level of this particular attribute and a $T = 40$ score indicates a fairly low level. However, an MMPI *T* score of 60 on scale 8 (Schizophrenia) does not indicate a relatively high degree of schizophrenia, nor does a *T* score of 40 indicate a low level. Thus, relatively normal profiles on the CPI are not only to be expected, but also can be interpreted successfully.

RELIABILITY AND VALIDITY

In general, the reliability and validity studies on the CPI compare favorably with those done on other personality inventories. Test-retest reliabilities for individual scales have ranged between a low of .51 for Flexibility to a high of .84 for Femininity/Masculinity. The overall median reliability was reported to be .68 (Gough, 1996). However, the retest interval was a long one year. Measures of internal consistency indicate considerable variability among the test items, but, overall, the scale constructions are adequate. Internal consistency ranged from a low of .43 for Masculinity/Femininity to a high of .85 for Well Being (median internal consistency was .76; Gough, 1996)

Factor analytic studies have been reported for both a two-factor and four-factor solution. In a general way, the factor structure suggests that elevations on CPI scales suggest personal adjustment whereas low scores indicate psychopathology (Higgins-Lee, 1990; Wallbrown & Jones, 1992). Megargee (1972) reported that the two factors of Internal Controls and Interpersonal Effectiveness accounted for a major portion of the variance on the original 1957 scales. Gough (1996) suggested five factors for the 1996 revision—Ascendence (dominance, empathy), Dependability (self-control, good impression), Conventionality (sociability, communality), Originality (flexibility), and Femininity/Masculinity. Factors I (Ascendance) and III (Dependability) roughly corresponded with the first two vectors or factor scorings that have been included in the 1987 and 1996 revisions. Vector 1 is a measure of introversion-extraversion and Vector 2 measures the extent to which a person is norm-favoring versus norm-doubting. An additional factor, Vector 3, provides an index of a person's psychological integration and self-realization.

The CPI has also been found to relate to most of the core five factors of personality (Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness). Both a conceptual and empirical analysis found that four of these five factors correlated highly with different clusters of CPI scales (Deniston & Ramanaiah,

1993; J. A. Johnson, 2000; McCrae, Costa, & Piedmont, 1993). For example, Openness to Experience was found to correlate most strongly with Achievement via Independence (.41), and Flexibility (.42,) but also with Capacity for Status (.38) and Social Presence (.42; McCrae et al., 1993). However, the Agreeableness factor was only minimally represented on the CPI (Deniston & Ramanaiah, 1993; McCrae et al., 1993). This is despite a rational analysis of the CPI scales, which would suggest that Agreeableness would be related to scales such as Socialization, Responsibility, Self-Control, and Good Impression. Despite this, the five-factor model was generally well represented on the CPI. This provides support that the CPI is measuring central aspects of personality and suggests the possibility that future CPI scoring might include an option for scoring these factors (Bolton, 1992).

In line with Gough's practical orientation, the main work on validation has been predictive (see Gough, 2000). Thus, the CPI has been less concerned with areas of psychometric elegance, such as whether the scales avoid overlap or measure psychometrically sound traits, than with the practical usefulness of the scales in providing accurate predictions. Specifically, persons scoring high on certain scales are more likely to be described in certain characteristic ways by those who know them. The CPI also focuses on predicting the types of things people do or say when placed in defined situations (i.e., leadership role). The scales themselves or the equations developed from various combinations of scales are able to predict a wide variety of different aspects of behavior. Many of the studies that have found useful levels of predictive validity are summarized later in this chapter in the Configural Interpretation section.

ASSETS AND LIMITATIONS

The CPI focuses on diagnosing and understanding interpersonal behavior in normal populations. Instead of focusing on pathology, it assesses areas such as self-control, dominance, and achievement. However, even though its emphasis is on assessing normal variations, extreme scores can also provide important information about the specifics of a person's expression of maladjustment, particularly with regard to interpersonal relationships (Cook, Young, Taylor, & Bedford, 1996; McAllister, 1996; Sarchione, Cuttler, Muchinsky, & Nelson-Gray, 1998). Whereas the MMPI is limited to use with primarily pathologically oriented populations, the CPI is appropriate for normal persons. Thus, it addresses issues that interest a great many people.

The main thrust of the research and construction of the CPI has been toward developing accurate, long- and short-term behavioral predictions. The focus is not so much on evaluating and predicting a specific, internal, unidimensional trait, but more on interpersonal behaviors and orientations. Gough (1968, p. 56) clarifies this by stressing that "a high score on a scale for social status does not mean that the individual has a 'trait' of high status; presumably, therefore, he may be already of high status, or possessed of those talents and dispositions that will lead him toward such attainment." Gough also stresses that certain interpersonal behaviors occur in specific contexts. For example, a person who scores high on dominance would be expected to assume control of a group requiring leadership. Thus, the longitudinal studies on the inventory have developed predictive strategies relating to areas such as police performance (Sarchione

et al., 1998), graduation from high school (Gough, 1966), grades in college (Gough & Lanning, 1986), choice of major field in college (Goldschmid, 1967), prediction of delinquent and criminal behavior (DeFrancesco & Taylor, 1993; Gough & Bradley, 1992), persistence among hospice volunteers (Lafer, 1989), and creative potential (Gough, 1992). A number of special scales have been developed for assessing specific areas. These are available through the CPI computer-scored report and include Managerial Potential, Work Orientation, Leadership, Social Maturity Index, Law Enforcement Orientation, and Creative Temperament (see Gough, 2000; McAllister, 1996; P. Meyer & Davis, 1992). The test has generally proven to be a useful tool in the area of prediction and, as a result, has been particularly helpful in counseling high school and college students as well as in personnel selection (Gough, 2000; P. Meyer & Davis, 1992).

Because the CPI's basic concepts were derived from day-to-day social interaction, it is relatively easily understood by a wide range of persons. Descriptions such as *dominant, achievement oriented,* and *self-controlled* are generally straightforward and are, therefore, not easily misinterpreted by untrained professionals. In contrast, providing feedback to clients who have taken the MMPI requires the clinician to rephrase psychiatric terminology into more approachable, easily understood terminology. Because the CPI relates to ongoing aspects of behavior, CPI interpretations are also likely to have more immediacy, relevancy, and impact on persons receiving feedback from their test results. These "folk concepts" also are generally found in all cultures and societies. Thus, Gough hoped that the inventory would have cross-cultural relevance and validity. Initial research does indicate that the CPI can be adapted to various cultures and that the concepts contained in the inventory do have cross-cultural relevance (Paunonen & Ashton, 1998). Although some cross-cultural research has been done on or using the CPI in different cultural contexts, more work still needs to be performed and subgroup norms should be applied when appropriate (Davis, Hoffman, & Nelson, 1990). Specific areas of future research should be the relationship between CPI scores and race, socioeconomic status, and other demographic variables. In particular, further research needs to be conducted on the ability of the CPI to predict relevant behaviors in a specific cultural context.

A number of predictive studies have been conducted from a research perspective, and several useful regression equations have been developed as aids in predicting behavior. However, extremely few studies have been performed to test the validity of predictions made by clinicians in actual practice (Gynther, 1978). It may be that clinical judgments based on the CPI are generally accurate, but, at this point, further empirical studies are needed for verification. It is something of a contradiction that a test with an emphasis on practical usefulness has not been sufficiently evaluated in the clinical context. Most predictive studies have attempted to estimate areas such as work outcomes, future college attendance, or grade point average in graduate programs. However, college attendance and grade point average do not necessarily correlate with later successful performance. For example, high medical school grades have not been found to correlate with later success as a physician (Loughmiller et al., 1970). This problem is certainly not unique to the CPI but is a general issue with many similar tests and relates to a difficulty in adequately establishing appropriate criterion measures. These issues suggest that test users should develop predictions based on test scores within limited and well-researched contexts. For example, if the CPI is being used to evaluate

prospective medical students, it should be made clear that predictions are useful only with regard to the students' academic performance and not to their overall clinical skills or later success as physicians.

One major past criticism directed toward the CPI is the lack of factor analysis in the development of the different scales (Eysenck, 1985). Factor analytic studies suggest that most of the variance can be accounted for by only two factors: interpersonal effectiveness and internal controls (Megargee, 1972). This conclusion is further supported in that many of the scales are highly correlated, are conceptually similar, and have extensive item overlap. Gough (1996, 2000) has responded by pointing out that the scales were designed to assess constructs that are in most people's minds on a daily basis. Any scale overlap, then, might accurately reflect the conceptual overlap in common folk concepts used on a daily basis, such as the self-control and high degree of socialization involved in responsible behavior. Even if many of the scales are quite similar, there is accumulating evidence that the scales measure what they were designed to measure. The lack of factor analysis is further corrected by inclusion in the 1987 and 1996 versions of three different factor-analytically derived scales that measure extraversion-introversion (externality-internality), norm-favoring versus norm-questioning, and degree of self-realization (Gough, 1996). In addition, more recent factor analytic studies indicate that four of the five core factors of personality are strongly represented in the CPI items and scales (McCrae et al., 1993; Walbrown & Jones, 1992).

A further limitation of the CPI is the insufficient number of studies undertaken on the meaning of pairs or triads of scales (Baucom, 1985). This may result partly from the formidable number of possible CPI code types (compared with the MMPI's more manageable 45 possible combinations). In addition, many persons score within a relatively narrow range, which makes configural interpretation more difficult because there are less likely to be clearly defined clusters of high and low scales (Shaw & Gynther, 1986). In contrast, extensive fruitful research has been conducted on two- and three-point codes for the MMPI. Some of the work conducted on CPI code types is summarized later in the Configural Interpretation section. Gough's more recent work on the three vectors (externality-internality, norm-favoring/norm questioning, and realization) has also provided information regarding composite subscale or factor scores (see Structural Scale Interpretations section). In addition, McAllister (1996) has listed 152 different combinations of scale scores based on a combination of empirical research, rational considerations, and clinical experience. However, more research needs to be done on the many possible two- and three-point codes that could potentially be derived from the CPI.

In developing accurate clinical interpretations from the CPI, it is essential to consider the implications of factors such as the overall life situation of the examinee. For example, the profile of a 15-year-old on the CPI scale for Psychological-Mindedness (*Py*) has a meaning different from that of a profile of a person of 55. Another important consideration is the purpose for which the person believes he or she is being examined. A person who is taking the test in a conscious effort to receive a discharge from the military will likely bias his or her responses in a direction different from a person seeking employment. It is also essential to look at overall patterns of scores rather than "single sign" indicators because corresponding elevations on other scales can elaborate or modify the meaning they have for one another. Thus, clinicians should always keep in mind

the implications of an examinee's overall life situation, age, education, perceived reason for assessment, and pattern of scores.

A further issue relates to the degree of comparability between the 1996 revision and previous versions. In the 1996 revision, the total number of items was reduced (from 462 to 434). However, high correlations ranging between .96 and 1.00 indicated sufficient comparability between the scales that previous research and interpretations based on earlier versions can be transferred to the 1996 (Form 434) version. Many of the prediction equations have been (and are being) updated. However, the new scales may be different in some yet-to-be defined ways. As research on the 1996 revision continues, this issue can be progressively addressed and interpretations altered (or reinforced).

The CPI, then, is an extremely useful test in the assessment of the interpersonal characteristics of relatively normal persons. It measures variables that interest a great number of people, providing helpful behavioral predictions, and uses routine, day-to-day interactional concepts. For these reasons, the CPI is extensively used in personnel selection and vocational guidance (Bolton, 1992; Gough, 2000; McAllister, 1996; Megargee, 2002; P. Meyer & Davis, 1992). Significant limitations and cautions relate to limited validity studies in clinical settings, few empirical studies on the meaning of two- and/or three-point elevations, and the unknown (but continually emerging) comparability between the 1996 revision and the previous versions.

INTERPRETATION PROCEDURES

The examiner should note the length of time it takes a person to complete the test. A person with an IQ within the normal range would be expected to complete the test in 40 to 60 minutes. A computerized administration of items can reduce the time from 25 to 45 minutes. If he or she takes 60 to 90 minutes or more, it suggests one of the following:

1. A major psychological disturbance such as severe depression or functional psychosis.
2. Obsessive concern with detail and/or indecisiveness.
3. A low IQ combined with poor reading ability.
4. Cerebral impairment.

Tests that are completed in 20 minutes or less suggest:

1. An invalid profile.
2. An impulsive personality.
3. Both 1 and 2.

An alternative form of administration is an oral or tape-recorded format, which would be particularly relevant for persons with unusually low reading skills. If time efficiency is important, a 250-item short form is currently being developed.

Scoring for the 1996 (Form 434) must be done by computer through Consulting Psychologists Press or through onsite scanning. Scoring is also performed for additional

special purpose scales (i.e., Managerial Potential, Work Orientation). Examiners may wish to extend the traditional scale information by using regression equations for areas such as high school achievement, parole success, or medical school performance (see Table 9.2).

The steps for interpreting the CPI are described in the following sections.

1. Determine Profile Validity

The CPI, similar to the MMPI, has built-in scales and relevant regression equations to detect invalid profiles. This is important because Gough (1987) has estimated that, in large-scale testing situations, approximately 1.7% of all profiles are invalid (.6% "fake good," .4% "fake bad," 0.7% random answering). As would be expected, invalid profiles may occur more frequently in other contexts such as the quite high 7.5% fake good rate noted among male applicants to become police officers (Gough, 1996).

An initial consideration in evaluating the profile validity is to note the number of items that have been left blank (available on the computer-generated profile). If 25 or more spaces are blank, the test results may not be valid. The examiner should also make sure the subject has not marked a large number of questions (25 or more) with both "true" and "false" on the same item. Yet, another area that should be checked is the possibility of random answering. The subject may appear to have answered randomly simply because he or she was out of step between the number's questions in the answer sheet and test booklet, or may answer randomly in an attempt to hide his or her poor reading ability. A good indicator for the possibility of random answering is a low score ($T = 30$ or less) on *Cm.*

"Faking bad" can usually be detected based on the presence of extremely low scores on Well Being (*Wb;* $T = 30$ or less) and Communality (*Cm;* $T = 30$ or less). A low score ($T = 40$ or less) on Good Impression (*Gi*) is also frequently associated with "faking bad." It should be stressed that a subject who "fakes bad" is not necessarily maladjusted. Rather, it indicates that the specifics of his or her disorder cannot be evaluated because of the distorting effects of the person's need to create an impression of the seriousness of his or her problem. Thus, it is important to assess why the person is "faking bad." It might, for example, represent a "cry for help" in which suicide is a serious possibility, or the person might be malingering because of numerous secondary gains.

To determine whether a subject is "faking good," the most important scale to evaluate is Good Impression (*Gi*). "Fake good" profiles usually have high scores ($T = 70$ or higher) on this scale. Usually, when a person is asked to "fake good," all the scales with positive social connotations are elevated but *Gi* is still relatively higher than the others. Sometimes it may be difficult to differentiate between someone who has a superior level of adjustment and a person who is "faking good." The most significant consideration in making this distinction is the person's history. An individual with a history of poor adjustment, combined with an unusually high *Gi,* is probably "faking good," whereas a person with a history of good adjustment and a moderately high *Gi* is probably expressing his or her superior level of adjustment.

These critical-scale values for the three validity scales can generally serve as clinical tools to detect invalid profiles. However, Gough (1996) notes that a significant

number of errors are still likely to occur. Instead, he recommends the following equations (*raw scores must be used in all equations*):

Fake good = 41.225 + .273 *Do* + .198 *Em* + .538 *Gi* − .255 *Wb* − .168 *Fx*

Fake bad = 86.613 − 1.000 *Cm* − .191 *Wb* + .203 *Ac* − .110 *Fx*

Random = 34.096 + .279 *Gi* + .201 *Wb* + .225 *Py* + .157 *Fx*

The following optimal cutoff scores for these equations should be used:

Fake good = 60.60 or greater

Fake bad = 59.50 or greater (and if the score on random is less than or equal to 48.01)

Random = 48.01 or less (and the score on fake bad is equal to or less than 59.50)

Research using simulators suggests that these equations detect fake bad protocols 84% of the time, fake good protocols 64% of the time, and random protocols about 87% of the time (Gough, 1996). It is hoped that clinicians would be able to increase these detection rates by considering the person's overall context—especially his or her reasons for taking the test and past history.

2. Note Vector Scale Patterns

A basic underlying description of core aspects of the person's functioning can be determined by interpreting the three vector scores of internality/externality, norm favoring/norm doubting, and level of realization. These provide a context for understanding other more specific aspects of personality (see Vector Scale Interpretation section).

3. Note General Level of Elevations/Lowerings

Scores of $T = 50$ or more usually suggest a positive area of adjustment. Scales that are well below $T = 50$ indicate specific problem areas. However, the clinician must also interpret these scores in the overall context of assessment, taking into account variables such as the person's age, occupational level, cultural background, and educational level. For example, a high school student with an Intellectual Efficiency (*Ie*) scale score of 60 represents a fairly high level of this characteristic, whereas the same score for a medical student represents a relatively low level when compared with his or her fellow students.

4. Note Patterns of Elevations/Lowerings on Different Clusters and Classes

After looking at possible areas of adjustment and maladjustment, the clinician can then further evaluate the profile by examining the average elevations on the different clusters or classes (Table 9.1) as organized by Gough (1996, 2000). For convenience, the

Table 9.1 Cluster analysis

Cluster	Meaning	Scales
1	Interpersonal style and orientation	Dominance (Do), Capacity for Status(Cs), Sociability (Sy), Social Presence (Sp), Self-Acceptance (Sa), Independence (In), Empathy (Em)
2	Normative orientation and values	Responsibility (Re), Socialization (So), Self-Control (Sc), Good Impression (Gi), Communality (Cm), Well Being (Wb), Tolerance (To)
3	Cognitive and intellectual function	Achievement via Conformance (Ac), Achievement via Independence (Ai), Intellectual Efficiency (Ie)
4	Role and personal style	Psychological-Mindedness (Py), Flexibility (Fx), Femininity/Masculinity (F/M)

clusters are separated on the profile sheets by black, vertical lines. If most or all of the scales in a particular cluster are clearly above $T = 50$, the qualities represented by the cluster are areas of strength. In contrast, scores well below $T = 50$ represent areas of difficulty.

The clusters listed in Table 9.1 are organized according to conceptual similarity rather than statistically derived categories. In contrast, Gough (1987, 1996) also recommends examining the scales based on five factors that have been statistically derived from more empirical relations. Factor 1 (*Do, Cs, Sy, Sp, Sa, In, Em*) indicates a person's level of social poise and interpersonal effectiveness. Factor 2 (*Wb, Re, So, Sc, To, Gi, Ac*) provides a general index of mental health, adjustment, and social conformity. The third factor (*Ai, Fx, To, Ie, Py*) includes scales that are characterized by assessing the extent to which a person can think and behave independently. The fourth factor is composed of scales *Cm, Re, So,* and *Wb* and measures the extent a person adheres to social norms and expectations. High scorers (all above $T = 50$) are likely to be conventional and place a high emphasis on doing and perceiving things correctly, whereas low scorers (all below $T = 50$) are more unconventional, individualistic, and likely to perceive the world in more unusual ways. The final, fifth factor is composed of Femininity/Masculinity and assesses a person's degree of aesthetic interests, dependency, and sensitivity. Clinicians can gain useful information by using either Gough's clusters (classes) or the more empirically derived five factors.

5. Evaluate the Meaning of the Scores on Each Individual Scale

Whereas the different clusters, factors, or vectors provide general impressions for certain areas of functioning, the clinician can obtain more specific information by evaluating

each scale individually. This involves looking at the relatively highest and lowest scales and developing a description of the dynamics involved with these scales. The meanings associated with specific high or low scores can be determined by considering the relevant scale descriptions in the Individual Scales section. The general personality descriptions and discussions of the scales have been adapted and modified from the publications of Gough (1975, 1996, 2000), McAllister (1996), and Megargee (in 2002). Additional relevant material has also been included and is cited accordingly. The short list of most frequently used adjectives (provided at the end of the sections on high and low scores) is based on ratings reported by Gough (1987, 1996). The adjectives were included based on their occurrence in two or more instances from within the different lists of ratings on the Adjective Check List made by peers, spouses, or CPI assessment staff.

6. Note Scale Configurations and Calculate Regression Equations

Initial hypotheses can be further evaluated by consulting the section in this chapter dealing with typical scale configurations for different areas, including intellectual level, achievement, leadership, adjustment, and specific syndromes. This evaluation may also involve calculating and interpreting the regression equations, which are included in the section on configural interpretations and summarized in Table 9.2. However, some caution should be used with them because they were derived from previous versions of the CPI and updated versions have not yet been reported.

Table 9.2 Summary of CPI equations used for making predictions

1. Achievement (High School) = 20.116 + .317 Re + 192 So – .309 Gi + .227 Ac + .280 Ai + .244 Ie
2. Achievement (High School – Using IQ) = .786 + .195 Re + .244 So – .130 Gi + .19 Ac + .179 Ai + .279 IQ
3. College Attendance = 17.822 + .333 Do + .539 Cs – .189 Gi + .740 Ac
4. Achievement (Introduction to Psychology) = 35.958 – .294 Sy – .180 Sp + .185 Re – .189 Sc – .152 Gi – .210 Cm + .275 Ac + .523 Ai + .241 Ie + .657 Py
5. Male GPA = .16 SAT (Math) + .11 So – .19 Sp + .17 Fe
6. Female GPA = .25 SAT (Verbal) – .14 Sp + .06 Re + .20 Ac + .08 Fe
7. GPA = 30.60 – .26 Wb + .35 Re – .19 Gi + .39 Ai + .22 Ie + .36 Py
8. Teaching Effectiveness = 14.743 + .334 So – .670 Gi + .997 Ac + .909 Py – .446 Fx
9. Medical Promise = .794 Sy + .602 To + 1.114 Cm – .696 Cs
10. Dental Performance = 29.938 – .110 Sp + .148 Re – .262 Gi + .727 Ac + .230 Py
11. Leadership (Social) = 14.130 + .372 Do + .696 Sa + .345 Wb – .133 Gi + .274 Ai
12. Parole Success = 45.078 – .353 Sp – .182 Sa + .532 So + .224 Sc
13. Social Maturity = 25.701 + .408 Re + .478 So – .296 Gi

7. Integrate Data into a Profile Description

The final step in interpretation is to integrate all the data into a profile description. Essential is the clinician's ability to assess the interactions between two or more scales. This suggests that, after a specific trend has been established, the clinician should elaborate on it by evaluating how the other scales change their meaning for the individual (cf. McAllister, 1996; Meyer & Davis, 1992). For example, dominance may be expressed in numerous ways, including rebellion, high achievement, leadership, or delinquency. When these elaborations have been made in the test data, the clinician can then seek outside confirmation through personal history, behavioral observations, and additional test data.

VECTOR SCALE INTERPRETATION

The major addition to the 1987 revision was the development and inclusion of three structural scales. These three scales form what Gough (1996, 2000) has referred to as the *cuboid model* because a person's position can be conceptualized as existing in three-dimensional space. Each of the dimensions was based on a factor analysis of the different items on the CPI. The first theme or factor that seemed to emerge referred to elements of extraversion, self-confidence, assertive self-assurance, and social poise. Items measuring these dimensions were formerly used to develop a scoring for the first vector (or structural scale), which Gough (1987, 1996) referred to as externality-internality. The second factor was related more to the degree to which a person accepted societal norms and included areas such as social conformity, personal integrity, self-control, and disciplined effectiveness. Scoring for these qualities was formally developed into a second vector, which Gough (1987, 1996) referred to as norm-favoring versus norm-questioning. The final, third vector was labeled realization and assesses the degree to which a respondent has developed a sense of self-realization and psychological integration.

On the CPI profile sheet, the first two vectors (externality-internality and norm-favoring versus norm-questioning) are combined to place a person into one of four specific types (Alpha, Beta, Gamma, Delta) based on the interaction between Vectors 1 and 2. The primary emphasis on structural scale interpretation is to understand the meaning associated with these four types. The third vector is used to provide additional meaning to these four types by considering the degree to which the person has managed to integrate them into a fully developed (self-realized) person. Vector 3 is rated on a scale between 1 and 7, where 1 represents no or little integration/realization and 7 represents an unusually high level. Gough (1987, 1996) describes these more specifically as 1 = poor, 2 = distinctly below average, 3 = below average, 4 = average, 5 = above average, 6 = distinctly above average, and 7 = superior.

Any interpretation of type should take into account both the extent to which the person has realized his or her type (Vector 3) as well as the relative strength with which he or she represents the type. For example, an Alpha combines qualities of extraversion and norm-favoring. These qualities would be far stronger if they scored quite high on both extraversion and norm-favoring (Vectors 1 and 2) than if they scored merely in the borderline areas. Specific interpretations and the implications of their degree of realization

are described in the following section and were derived from descriptions provided by Gough (1996, 2000) and McAllister (1996).

Alphas

Persons scoring in this quadrant tend to be highly extraverted and to adhere to societal norms. They will be good leaders because they are task-focused and productive but also interested in associating with others. Their social style may be somewhat managerial. Externally, they may be assertive, talkative, and have high levels of achievement and social presence. If highly realized (note Vector 3), Alphas may be charismatic leaders and help to create social change. If undeveloped, they might become manipulative, self-centered, and concerned only with achieving their own ends regardless of consequences to others.

Betas

Betas combine qualities of both introversion and norm-favoring. Thus, they prefer external structure and are generally most comfortable in the role of a follower. They have a high degree of self-control, are highly dependable, conservative, value traditions, and may place the needs of others before their own. If highly realized, they can be nurturant, represent ideal models of goodness, and convey conventional sources of wisdom. Poorly developed Betas might be nonresponsive, overly conformist, inflexible, constricted, and rigid.

Gammas

Gammas are extraverted and, at the same time, question traditional beliefs and values. Thus, they make their questions, beliefs, and challenges quite apparent. These are the skeptics, doubters, and persons who might try to change society. They perceive the world in highly individualistic ways but are still actively involved with others. Often, they might try to test limitations imposed on them and do so in a rebellious, self-dramatizing manner. At their best, Gammas are innovative, visionary, perceptive, and imaginative. They are likely to be inventors, create new ideas, and push their field to new limits. If inadequately developed, they are intolerant, belligerent, self-indulgent, rebellious, and disruptive.

Deltas

Persons scoring in this quadrant have qualities of introversion and also question traditional values and beliefs. As a result, Deltas are highly reflective, somewhat detached, preoccupied, and possibly overly absorbed in their own fantasies and daydreams. They might prefer that others make decisions for them and, if extreme, may live primarily in their own private world. If fully developed, they might be highly imaginative, artistic, visionary, and innovative. However, they run the risk that their innovations may go unnoticed because they rarely make a production of their activities. If poorly developed, Deltas may be poorly organized, withdrawn, aloof, self-defeating, and at risk of decompensating.

INDIVIDUAL SCALES

1. Dominance (*Do*)

The *Do* scale measures areas of leadership ability and has become one of the most validated scales on the CPI. It includes verbal fluency, persuasiveness, and the extent to which a person is likely to take charge of a situation. Thus, high scorers are persistent in approaching a task and usually take the initiative in interpersonal relationships. However, this description is more characteristic of the style in which high-scoring males express their dominance. High-scoring females express their dominance either by initiating attempts to choose a leader, or by being somewhat coercive, aggressive, or impatient. The contents of the items deal with social poise, confidence, verbal fluency, persuasiveness, and a sense of duty.

It should be stressed that the conditions in which leadership occurs are at least as important as the actual trait. This means that, when a situation arises requiring leadership, high scorers usually become leaders rather than followers. More specifically, they are more likely to be the ones to set limits, and become more assertive, goal oriented, and clear and direct regarding their requests. They adopt this role relatively comfortably and naturally. In contrast, low scorers experience discomfort when requested to take charge. They may be either more submissive, in which case they prefer others to control and direct them, or merely socially isolated and introverted, in which case they do not want to control others but also do not want others to control them. They may even actively resist efforts that are made to control them.

High Do *(*T = *65 or More)*

High scorers on *Do* are strong in expressing their opinions and in reaching their goals. This may range from being highly assertive, in which they are clear and direct in expressing their needs, to being aggressive, in which they are more forceful. They would rather take charge of a situation and can effectively do so because they have excellent abilities to plan and are self-confident when directing others. Persons high in dominance can use and develop the resources available to them and often express a sense of optimism. They are generally able to define their goals and work persistently to attain them. They would not be particularly compromising nor would they be the type of person to whom others would feel comfortable admitting their weaknesses. The most frequent adjectives used to describe them are assertive, confident, dominant, task-oriented.

Moderate Do *(*T = *50 to 65)*

Moderate *Do* scale scorers have the capacity for leadership but do not, under ordinary circumstances, seek opportunities to use this ability.

Moderately Low Do *(*T = *40 to 50)*

With moderately low *Do,* people usually feel uncomfortable when leadership is required and much prefer being in the follower role. They are participants rather than organizers. Although some persons who are low in dominance are effective in relatively high leadership positions, they are uncomfortable with this aspect of their job, and usually have a democratic and participative style of decision making. However,

most persons scoring low on *Do* experience a difficult time planning and, as a result, may sometimes appear reckless and impulsive. They are likely to believe and adhere to the beliefs of others and can, therefore, be easily influenced. Often, they have a difficult time making direct requests, and are usually seen as nonassertive. Low scorers, particularly females, are seen as submissive, shy, timid, and inhibited.

Low Do *(*T = *40 or Less)*

Extremely low scores on *Do* suggest a general pattern of maladjustment They are likely to be socially withdrawn, insecure, and shy. They see themselves as having little or no leadership ability and dislike being directly responsible for either their own actions or the actions of others. They may be passive, require prodding, and attempt to avoid situations that are likely to produce tension and pressure. The most frequently used adjectives to describe these persons are shy, timid, submissive, withdrawn, quiet, retiring, unassuming, silent, and inhibited.

2. Capacity for Status (*Cs*)

An individual's capacity for status has been defined by Gough (1968, p. 61) as equal to the "relative level of income, education, prestige, and power attained in [his or her] social-cultural milieu." This definition focuses on status as it has been achieved, but the *Cs* scale looks at status more as a trait associated with features such as ambition and self-assurance. The specific trait of capacity for status suggests that, eventually, a person will achieve and maintain a position of status. Thus, in creating the scale, Gough looked at the specific trait variables that would eventually lead to a higher status position. These traits include perseverance, self-direction, ambition, and self-confidence. Persons seeking status are usually willing to go through a fairly high degree of discomfort and personal change to achieve their goals. In the scale construction, there is some overlap of test items with Social Presence (*Sp*), Intellectual Efficiency (*Ie*), and Self-Acceptance (*Sa*), indicating that capacity for status also includes dimensions of social poise, efficiency, and self-confidence. The item content also reflects an absence of fears or anxieties, a high degree of social conscience, an interest in belonging to various groups, and an interest in literary and aesthetic activities.

High Cs *(*T = *60 or More)*

Individuals with high scores on *Cs* are characterized as independent, imaginative, and will take advantage of opportunities that are presented to them. They are highly self-directed, achievement-oriented, and able to respond to their environment in a manner designed to further their own goals. Their aspirations are high, and they have excellent verbal fluency. Extremely high scores suggest they are overbearing, arrogant, aristocratic, and feel superior. The most frequent adjectives used to describe these high scorers are ambitious, confident, intelligent, versatile, enterprising, interests wide, assertive, and (having) initiative.

Moderate Cs *(*T = *45 to 60)*

As might be expected, moderate scorers are somewhat goal oriented and relatively highly motivated to achieve. They are willing to change and adapt their lives to a

certain extent in their attempts to achieve status. They are also moderately ambitious and self-assured.

Moderately Low Cs *(*T = *35 to 45)*

These individuals are minimally goal oriented, but their general lack of self-direction is not sufficiently low to impair their level of functioning. They are unwilling to make many personal sacrifices to achieve power, prestige, or a higher income. Accordingly, they are likely to experience considerable indecisions regarding their careers (J. L. Newman, Gray, & Fuqua, 1998).

Low Cs *(*T = *35 or Less)*

Persons who score extremely low on *Cs* usually have a low level of energy and are relatively rigid and inflexible. Their interests are extremely narrow, and they are likely to have little curiosity about their environment. They are usually resentful of their current position, which results in tension, restlessness, and depression. In the face of difficulties, they usually give up easily and withdraw. Their thinking is commonplace, unimaginative, literal, and slow. The most frequent adjectives used to describe them are shy, timid, silent, interests narrow, quiet, and simple.

3. Sociability (*Sy*)

The Sociability scale was originally designed to measure the extent to which a person participates in social activities. It was later generalized to differentiate between a person who is outgoing, extraverted, and sociable versus one who is more introverted, withdrawn, and prone to avoid social visibility. There is a great deal of item overlap with Intellectual Efficiency (*Ie*), Social Presence (*Sp*), Self-Acceptance (*Sa*), and, to a much lesser extent, Achievement via Independence (*Ai*), Dominance (*Do*), Capacity for Status (*Cs*), and Achievement via Conformance (*Ac*). The questions deal with enjoyment of social interactions, a sense of poise, self-assurance in dealing with others, and interest in cultural and intellectual activities.

High Sy *(*T = *60 or More)*

High scorers on *Sy* have some of the same traits as persons scoring high on Capacity for Status (*Cs*), such as a greater sense of maturity and a wide range of interests. They are also described as outgoing, sociable, and confident. In general, they feel comfortable in social settings and can easily mix with others. They feel comfortable around large groups of people and dislike working alone. They have well-developed social skills and generally make a good impression. The most frequently used adjectives to describe them are outgoing, sociable, confident, ambitious, aggressive, energetic, talkative, assertive, and enterprising.

Moderate Sy *(*T = *50 to 60)*

Persons in this range have an average level of extraversion and are relatively comfortable in most social situations. Although they prefer to be around others, they do not, by any means, exclusively orient their lives in this direction.

Moderately Low Sy *(*T = *35 to 50)*

Such persons are able to interact with groups of people without experiencing an excessive amount of discomfort, but they prefer to be alone. They feel somewhat anxious around strangers and strongly prefer to be with persons with whom they are already acquainted. Usually, they dislike being the center of attention.

Low Sy *(*T = *35 or Less)*

Persons who score this low have a definite sense of awkwardness in social situations and frequently have bitter complaints about their lives. They have a marked lack of confidence in their social skills and, as a result, avoid most social encounters, especially in unfamiliar settings or with those they do not know. They might act in self-defeating ways, frequently perceive themselves as underachievers, and are prone to anxiety. The most frequently used adjectives to describe them are withdrawn, shy, retiring, quiet, timid, meek, quitting, reserved, and awkward.

4. Social Presence (*Sp*)

The Social Presence scale was intended to serve as a measure of a person's degree of poise, self-confidence, verve, and spontaneity in social interactions. It especially assesses the extent to which the person is self-assured and assertive. *Sp* is very similar to sociability in that an individual scoring high on *Sp* is outgoing, extraverted, and enjoys being around other people. However, a person who is sociable does not necessarily also have social presence even though this is often the case. Social presence implies not only that the person is sociable, but also that he or she has more of a need to have impact on others and is thus likely to be more verbally aggressive, irritable, and sarcastic. A person exerting social presence might manipulate and control others, especially by working on another person's defenses and self-deceptions. There is some overlap of items with Sociability (*Sy*), Self-Acceptance (*Sa*), and, to a lesser extent, Capacity for Status (*Cs*) and Intellectual Efficiency (*Ie*). The primary content of the questions relates to a person's poise and the degree to which he or she enjoys social interactions.

High Sp *(*T = *65 or More)*

High scorers are often described as being unconventional, spontaneous, witty, and perceptive. They are usually concerned with their own pleasure in interpersonal relationships and often manipulate interactions to feel a sense of personal power. Thus, they not only like to be with other people, but also want to be in control. Their expression of ideas and vocabulary is excellent, as are their social skills. They are often perceived as imaginative, socially relaxed, and generally make a good impression. Extremely high scorers might be manipulative, highly energetic, and feel offended if people do not pay attention to them. The most frequently used adjectives to describe them are outgoing, confident, versatile, talkative, and adventurous.

Low Sp *(*T = *40 or Less)*

Whereas high scorers are unconventional and uninhibited, low scorers are extremely cautious and concerned with proper etiquette. They feel that others should conform to set, predefined standards and are disapproving of nonconforming behavior. Their view

of what is correct and incorrect falls within relatively narrow limits. In their relationships with others, they emphasize cooperation rather than manipulation and are likely to be kind, appreciative, patient, and serious. However, this kindness and appreciation are expressed only when the behavior of others falls within their definition of *conventional.* They would most likely feel anxious when expected to alter their routine. They are moralistic regarding the behavior of others, but also can be made to feel guilty regarding their own behavior. Extremely low scorers might lack energy, avoid being the center of attention, and feel uncomfortable when required to use their influence on others. The most frequently used adjectives to describe them are shy, withdrawn, retiring, silent, quiet, timid, and inhibited.

5. Self-Acceptance (*Sa*)

The Self-Acceptance scale was intended to assess factors such as personal worth, self-acceptance, and capacity for independent thinking and action. Furthermore, it was hoped that *Sa* could "identify individuals who would manifest a comfortable and imperturbable sense of personal worth, and who would be seen as secure and sure of themselves whether active or inactive in social behavior" (Gough, 1987, p. 10). Although persons scoring high on *Sa* would be less likely to become upset, the *Sa* scale should not be used as an index of adjustment and is not related to the absence or presence of pathology. For example, a person might be high in self-acceptance, yet still be rebellious, impulsive, and generally indulge in antisocial behavior. In fact, persons scoring extremely high on *Sa* are quite likely to be egocentric and indifferent, sometimes even to the point of narcissism. The scale questions have some overlap with Sociability (*Sy*), Social Presence (*Sp*), and, to a lesser extent, Capacity for Status (*Cs*). There is some negative overlap in which answers are scored in the opposite direction from Capacity for Status (*Cs*). Thus, a number of statements deal with social poise and self-confidence. Additional areas of item content relate to an accepting attitude toward social prohibitions, attention to duty, consideration of others, and an acceptance of human frailties.

High Sa *(T = 65 or More)*

Individuals scoring high on *Sa* are comfortable with themselves, self-reliant and independent, and are usually polished, sophisticated, enterprising, and self-seeking in social relations. They also have a clear sense of self-definition, and are characterized as being self-confident and outgoing. However, *Sa* should not necessarily be tied with sociability because self-acceptance can be high regardless of the quantity of interaction with others. The scale is slightly correlated with hypomania, which has often been formulated as a defense against depression. Thus, extremely high scores may suggest an inflated sense of self-acceptance with underlying, but unacknowledged, feelings of self-criticism, pessimism, and hopelessness. The most frequently used adjectives to describe high scorers are outgoing, self-confident, talkative, ambitious, and assertive.

Moderate Sa *(T = 50 to 65)*

These persons have an average or somewhat above-average level of confidence, with a generally good sense of harmony and internal balance. They are somewhat adventurous and outgoing.

Moderately Low Sa *(*T = *35 to 50)*

Moderately low scorers are somewhat low in self-confidence and have some significant doubts about themselves. For the most part, they can adequately cope with their lives, but they are prone to periods of insecurity and depression. One way in which they often attempt to adapt is through conformity and conventionality, which frequently has the desired effect of making their world safer and more predictable.

Low Sa *(*T = *35 or Less)*

Such individuals have a pronounced lack of self-confidence. They are usually described as ordinary and have "flat" or unidimensional personalities. They achieve a moderate degree of safety in their world by withdrawing, quitting, and maintaining a relatively narrow range of interests. They are likely to have a strong sense of insecurity, are afraid to take risks, and have low levels of self-confidence. Although usually submissive and conventional, they may at times impulsively act out in a reckless manner, almost as a form of rebellion against their largely self-imposed conventionality. The most frequently used adjectives to describe them are shy, withdrawn, retiring, silent, quiet, timid, and inhibited.

6. Independence (*In*)

The Independence scale measures the extent to which a person strives toward vocational and interpersonal autonomy. Conceptually, it overlaps with *Ai* in that they both assess the value a person places on working away from the restrictions, expectations, and influence of others. It also has similarities with *Sa* because persons high in both *In* and *Sa* are likely to be self-assured and self-reliant. Similarities can also be found with *Sp* (both witty, animated) and *Do* (both like to be in control).

High In *(*T = *65 or More)*

Persons scoring high on *In* are self-assured, confident, and possess social presence. Their vocabulary is likely to be wide, and they are intelligent, self-reliant, witty, animated and, as a result, are likely to make a good impression. However, they are not necessarily affiliative and friendly. They are perceived as resourceful, confident, self-sufficient, and capable. If they believe in a concept, position, or fact, they will defend it without bending to external pressure. Interpersonally, they are likely to be dominant; vocationally, they have high needs for achievement, and they are willing to work to achieve higher status. They also tend to be morally responsible and have high levels of self-control. They are most frequently described as confident, independent, aggressive, (having) initiative, and assertive.

Moderately High In *(*T = *55 to 65)*

Moderately high scorers have many of the previous characteristics in that they are confident, goal-oriented, and able to rely on their own evaluations and directions. They are assertive and usually can deal effectively with others.

Low In *(*T = *30 to 45)*

Low scorers need to rely on others for decisions and directions. They are likely to avoid conflict, competition, and experience discomfort when having to assert themselves.

Assets include an excellent ability to cooperate and blend with the requirements and needs of others.

Extremely Low In *(*T = *30 or Less)*

If persons score in the extremely low range, it suggests they are dependent and lack self-confidence. They will probably accept domination from others, partially because they feel uncomfortable having to face uncertainty. Often they experience worry and anxiety and are reluctant to express their own ideas. Assets include tolerance, adaptability, generosity, and helpfulness. Frequent adjectives used to describe them are timid, shy, cautious, meek, submissive, unassuming, and nervous.

7. Empathy (*Em*)

The Empathy scale attempts to measure the degree to which a person perceives and can feel the inner experience of others. It also measures related abilities, including social skills, confidence, social presence, leadership, and extraversion. The major underlying themes to the scale are that "empathic persons are characterized by a patient and forbearing nature, by affiliative but socially ascendant tendencies, and by liberal and humanistic political and religious attitudes" (Greif & Hogan, 1973, p. 284).

High Em *(*T = *65 or Above)*

High scorers are intuitive, perceptive, verbally fluent, have a wide range of interests, and are usually perceived by others as interesting. In addition, they are highly creative, spontaneous, and able to use their imagination in a number of areas. They have social presence, and are animated, witty, and make a good impression. Thus, they are interpersonally effective, independent, and flexible. They are most frequently described as sociable, outgoing, versatile, spontaneous, interests wide, confident, and humorous.

Moderately High Em *(*T = *55 to 65)*

Moderately high scorers have some insight into the feelings and motives of others and are friendly, adaptable, and comfortable to be around.

Moderately Low Em *(*T = *30 to 45)*

Persons with moderately low scores are typically slow to understand the feelings and motives of others. They are perceived as having narrow interests and as shy, withdrawn, narrow-minded, and conventional.

Very Low Em *(*T = *30 or Lower)*

Extremely low scorers often feel bewilderment regarding the reasons others behave as they do. These individuals can often be insensitive and inconsiderate. Often they are shy, rigid, unfriendly, and others find it difficult to please them. They are uncomfortable with uncertainty and, as a result, might cling to a rigid set of morals and narrow range of behaviors, often becoming authoritarian and ethnocentric. Their fathers were probably distant, cold, and taciturn. The most frequent adjectives used to describe them are shy, silent, interests narrow, and conservative.

8. Responsibility (*Re*)

The intent of the *Re* scale is to assess the degree to which persons are "conscientious, responsible, dependable, [and] articulate regarding rules and order, and who believe life should be governed by reason" (Gough, 1968, p. 65). Although responsibility is somewhat related to sociability and self-control, it also stresses that values and controls are well-defined and significant factors in a person's life. The person who is highly responsible will sacrifice his or her own needs for the benefit of the group. Such people accept the consequences of their behavior, are dependable and trustworthy, and have a sense of obligation to the larger social structure. They are not necessarily leaders, but they do have a high sense of integrity and are committed to follow through on agreements they have made with others. In general, persons who express antisocial behavior score low on *Re,* whereas average or above-average scores are obtained by occupational groups in which responsible behavior and "attention to duty" are required. The *Re* scale is scored positively for items that reflect a high degree of commitment to social, civic, or moral values.

High Re *(*T = *60 or More)*

High scorers respond well to tasks in which they are required to be conscientious, dependable, and reasonable. They will give up their own personal satisfactions for the sake of the group and will honor any commitments they have made. Their approach to problem solving is extremely rational and clear. Usually, they have strong religious beliefs, a clear sense of ethics, and are concerned with philosophical issues. Their work is productive because their aspiration levels are high and their work style is dependable and responsible. Their behavior is courteous, polite, alert, energetic, honest, and direct. If given a choice, they will seek additional information to reduce risk and generally prefer to avoid risky behaviors themselves. The most frequent adjectives used to describe them are conscientious, responsible, dependable, thorough, industrious, and efficient.

Moderate Re *(*T = *40 to 60)*

Such persons respond well to tasks in which they are required to be conscientious, dependable, and reasonable. Generally, they are not comfortable taking responsibility for the behavior of others, but they are seen by others as reasonably conscientious and straightforward.

Low Re *(*T = *40 or Less)*

Individuals with scores this low show a lack of discipline, and are usually rebellious and impulsive. They have difficulty budgeting their finances and are seen by others as restless and careless. Their perceptions are tied to their own personal biases, and they are mainly concerned with their own needs. They often behave in exploitive and immature ways. Their histories usually reveal they had their first sexual encounters at an early age, they were underachievers in high school, had considerable disagreements with their parents, engaged in borderline delinquent behavior, and that often their fathers were alcoholics. Their external behavior is typically crude, unpredictable, rebellious, nonconforming, and self-indulgent. Internally, they feel dissatisfied, moody,

cynical, and distrustful. The most frequent adjectives used to describe them are rebellious, reckless, and pleasure seeking.

9. Socialization (*So*)

The Socialization scale was originally called the "Delinquency scale," and, as the name suggests, its intent was to assess the likelihood of antisocial behavior. The scoring was later reversed, its name changed, and it gradually became a measure of an individual's social maturity, integrity, and rectitude. It is probably Gough's favorite scale and is based on his theory that antisocial behavior is the result of a role that certain individuals assume. There has been an extensive accumulation of literature on this scale, due at least in part to Gough's personal interest in it. The research indicates that the *So* scale has excellent concurrent, predictive, and cross-cultural validity, and is probably the most validated and most powerful scale on the CPI (J. Collins & Griffin, 1998).

The Socialization scale was designed to measure the degree to which social norms are accepted and adhered to. An individual, then, can score on a continuum from extremely well socialized to highly antisocial. The scale also estimates the probability that a person will engage in behavior considered incorrect in his or her culture (J. Collins & Bagozzi, 1999). For example, the *So* scale has been able with relative accuracy to differentiate cheaters from noncheaters in a college population (Kipnis, 1968), and low *So* scores were related to a diagnosis of personality disorder (Kadden, Litt, Donovan, & Cooney, 1996; Standage, 1986, 1990). In addition, low scores were able to predict criminals who would reoffend (DeFrancesco & Taylor, 1993) and differentiate between delinquents and nondelinquents (Gough & Bradley, 1992). Schizophrenics making violent suicide attempts were found to have particularly low scores on *So* (Seeman, Yesavage, & Widrow, 1985). The *So* scale has also differentiated high school dropouts from graduates (Gough, 1966; Hase & Goldberg, 1967). In a further study, Wernick (1955) demonstrated that 50% of the low scorers who were hired as temporary Christmas help stole from the store and none proved to be satisfactory workers. Several researchers have found a negative correlation between *So* scores and a past lack of family cohesiveness, poor quality of parental care (Kadden et al., 1996; Standage, 1986, 1990), and physical abuse (Barnett & Hamberger, 1992). In addition, low scorers reported having a dysphoric mood (Kadden et al., 1996; Standage, 1990) and are more likely to experience career indecision (F. L. Newman, Ciarlo, & Carpenter, 1999). Thus, many items included in the scale are designed to determine whether the examinees experienced warmth and satisfaction in their family relationships. Some of the items also reflect the presence or absence of pessimism regarding a person's life and environment. The content of several other questions centers on whether examinees can properly evaluate the effects of their behavior as well as the extent to which they can be empathetic and sensitive to the feelings of others.

High So *(T = 65 or More)*

Persons scoring high on *So* are organized, adaptable, and efficient. They are highly dependable, but maintain this level of dependability by being cautious, self-controlled, and inhibited. In general, they are willing to trust others and express a fairly high level

of optimism. They are often described as kind, honest, and practical, and they typically come from a stable, cohesive family environment where warmth and concern were freely expressed. Often, they were overprotected, and their current behavior is usually relatively conventional. Their external behavior is typically gentle, considerate, honest, tactful, well organized, capable, and productive. Their values are conservative, and, as a result, they behave in an ethically consistent manner. Internally, they feel optimistic, stable, and well controlled. They are most frequently described as reliable, organized, dependable, stable, and cooperative.

Moderate So *(*T = *50 to 65)*

Individuals who score in this range are able to trust others and are generally accepting of the mores and rules established by society. They also tend to be inhibited and conventional, sometimes to the point of being overadapted, but not as much as those with higher *So* scores.

Moderately Low So *(*T = *30 to 45)*

Individuals who score in the lower ranges of *So* are somewhat impulsive and unreliable, and often have a difficult time trusting others. They are not usually followers; rather, they frequently question the rules given to them and, in general, do not have a high degree of respect for society's prescribed forms of behavior. They often express a moderate level of rebelliousness.

Low So *(*T = *30 or Less)*

Such persons have a far greater likelihood of antisocial behavior and are usually unreliable, unconventional, rude, defensive, and impulsive. They reject past family ties, primarily because their past family lives were filled with chaos and were unsatisfying. They were unhappy at home, experienced considerable friction with their parents, and were underachievers and sexually precocious. They experience a deep sense of alienation and have an extremely difficult time trusting people. They are likely to report having dysphoric moods. Scores below 25 are associated with personality disorder, especially borderlines and antisocial personality. Others see them as headstrong, unpredictable, deceitful, rebellious, and pleasure seeking. Internally, they feel cynical, moody, and often feel that their lives are meaningless. The most frequent adjectives used to describe them are reckless, impulsive, rebellious, unconventional, bitter, restless, and suspicious.

10. Self-Control (*Sc*)

The original intent of the *Sc* scale was to measure the degree to which a person can self-direct his or her own behavior. More specifically, high scores suggest that a person can delay his or her behavior and redirect it in a clear, goal-oriented manner. Thus, a certain degree of similarity exists between self-control and both responsibility and socialization. Gough (in Megargee, 1972) clarifies these concepts by stating that responsibility reflects the "degree to which controls are understood," socialization measures the "extent to which they influence a person's behavior," and self-control assesses the "degree to which the individual approves of and espouses such regulatory dispositions"

(pp. 65–66). Persons scoring high on *Sc* are self-directed, inhibited, and withhold their expressions of emotions and behavior. Some types of persons who score extremely high on *Sc* are often overcontrolled to the extent that, for short periods of time, they lose control and become explosive (Megargee, 1966; Megargee, Cook, & Mendelsohn, 1967). Individuals with low scores are impulsive and pleasure seeking, have difficulty delaying their impulses, and are not good at evaluating the consequences of their behavior. Thus, both extremely high and extremely low scorers are similar in that they have significant issues dealing with the management of impulses; however, they use opposite strategies to cope with these impulses.

The primary overlap of items for *Sc* is with *Gi,* and several items are also scored in a direction opposite from *Sp* and *Sa.* Some of the most important items emphasize that thought and rationality are the primary determinants of behavior. Furthermore, high scorers usually endorse items that indicate they take precautions to avoid irrational behavior and are generally socially inhibited.

High Sc *(T = 60 or More)*

Persons who score high on *Sc* are considerate, self-denying, and dependable. They have a high need for precision and make every attempt to be reasonable. Other people perceive them as considerate, wholesome, and dependable, but also as stubborn, rigid, and overconforming. They avoid situations in which they might be tempted to act impulsively, and are generally inhibited, lacking in spontaneity, and move slowly. Externally, their behavior is well organized, patient, capable, and fastidious. They are conservative and moralistic and behave in an ethical, conscientious, and consistent manner. Internally, they usually feel optimistic but serious. The most frequent adjectives used to describe them are moderate, calm, quiet, conservative, conventional, and conscientious.

Moderate Sc *(T = 45 to 60)*

Such persons are fairly conventional and somewhat inhibited. They carefully consider the consequences of their behavior before acting. Others usually see them as reasonable and dependable, although somewhat lacking in spontaneity.

Moderately Low Sc *(T = 30 to 45)*

Persons scoring in this range sometimes act in a spontaneous, impulsive manner but can usually delay their behavior. Thus, their level of impulsiveness is insufficient to impair their interpersonal and work relationships.

Low Sc *(T = 30 or Less)*

Low scorers have a marked difficulty delaying their behavior, are hasty in making decisions, and are usually individualistic and self-seeking. Their impulsiveness may sometimes cause tension in group activities, and they often regret having acted in inappropriate ways. At times, they can seem extremely unrealistic and headstrong. They are prone to develop relationships quickly, which often readily become chaotic and confused. The background of these individuals usually reveals they were sexually precocious, experienced considerable conflicts with their parents, and, academically, they were both underachievers and unhappy. Their external behavior is restless, excited, outgoing, rebellious, unpredictable, and they frequently perceive situations in sexual

terms. The most frequent adjectives used to describe them are impulsive, mischievous, restless, humorous, pleasure seeking, and adventurous.

11. Good Impression (*Gi*)

Although *Gi* is mainly a validity scale designed to detect persons who are "faking good," it also reflects the degree to which a person with a valid profile is concerned with creating a favorable impression on others. There is a fairly high degree of item overlap with Self-Control (*Sc*), which suggests that an important component of creating a favorable impression is a good ability to delay impulses. Also, a number of items make fairly obvious statements concerning the person's level of functioning, amount of antisocial behavior, the extent to which he or she is goal oriented, and whether he or she has complaints regarding personal failings. High scorers are prone to exaggerate their positive points and minimize their negative qualities. Furthermore, they state that they have a high level of confidence and self-assurance, and minimize anxieties or insecurities. They emphasize that they can adapt well to stress and that they have a stable personality. Finally, there are several items related to the extent to which individuals behave in a socially approved manner and experience harmonious relationships with others.

The *Gi* scale has generally been successful in detecting invalid profiles. For example, Dicken (1960), by using a cutoff score of $T = 60$, was able to detect in 79% of the cases the profiles of persons attempting to make a favorable impression. With somewhat different criteria, only 3% of a total sample of profiles of mixed "normal" and "fake good" were incorrectly classified. In the same study, Dicken also demonstrated that even though persons were, in some of the cases, attempting to "fake good" on other scales, *Gi* still showed the greatest increase. The practical importance is that *Gi* would still be expected to increase, even though a person might be attempting, for example, to exaggerate his or her level of responsibility. Thus, the use of *Gi* as a validity scale is not restricted to persons attempting to create a favorable impression in a global manner; it can be used to detect persons attempting to "fake good" along other specific dimensions as well. More precise and accurate classifications can be derived by using the equations summarized in Gough (1996) and included in the previous Determine Profile Validity section.

High Gi *(*T = *60 or More)*

An examinee's personal history provides the best guide for determining whether a score in this range reflects a "fake good" profile or is more likely to indicate a person with an excellent level of adjustment. For example, an alcoholic with a high *Gi* is probably either consciously attempting to create a favorable impression or demonstrating the use of denial, which is often associated with that disorder. A further possibility for an extremely high *Gi* is that the person may be unaware of the impression he or she creates on others and has an inflated self-image based on rigidly selective perceptions. The self-image of such persons would then be likely to be maintained by ignoring the feedback they receive from others and manipulating others to agree with the perceptions they have of themselves. They may be people-pleasers who will do anything to fit in and, as a result, are probably liked but not respected by others. Gough (1987) recommends that the ideal cutoff score for detecting a "fake good" profile is $T = 69$ for males and $T = 71$ for females.

If the profile is only moderately high and has been determined to be valid, then the person is likely to be conventional, adaptable, self-denying, and capable of a high degree of empathy. These people are often oversensitive to the criticisms of others, and usually respond by attempting to change and adapt to gain approval. They feel it is important to please others and to be seen in a favorable light. Others usually see them as kind, warm, considerate, and patient. They will probably attempt to overcontrol their needs and be moralistic, but will try to adapt by becoming considerate and tactful. The most frequent adjectives used to describe them are calm, conventional, conservative, and moderate.

Moderate Gi *(*T = *45 to 60)*

Persons with a moderate score on *Gi* are usually unselfish and concerned with making a favorable impression. They are able to take feedback from others and use it in a constructive way. Others perceive them as peaceable, trusting, understanding, and highly concerned with living up to their social responsibilities.

Moderately Low Gi *(*T = *30 to 45)*

Moderately low scorers are only minimally concerned with the impression they have on others to the extent that they are sometimes seen as insensitive. They feel that they alone are the judges of their behavior and, thus, rarely listen to the evaluations of others. They are often described as independent, witty, and, occasionally, temperamental and sarcastic.

Low Gi *(*T = *30 or Less)*

Persons scoring in this range are typically arrogant and actively reject the judgments of others. They are even prone to exaggerate their negative behavior in a rebellious way, and then expect this behavior to be tolerated and even accepted. Others describe them as temperamental, cynical, sarcastic, and overly frank to the point of being disagreeable. This usually has the effect of disrupting their interpersonal relationships. Their external behavior is typically rebellious, undiplomatic, critical, nonconforming, unpredictable, and self-indulgent. They might come from conflict-ridden families in which their mothers were nervous and dissatisfied. They are often perceived as insensitive and lack qualities of nurturance. Internally, they may often feel cynical, distrustful, and dissatisfied. Frequent adjectives used to describe them are temperamental, restless, and rebellious. Scores of $T = 35$ or less suggest a "fake bad" profile.

12. Communality (*Cm*)

The *Cm* scale is a validity scale originally designed to detect random answering. The questions are keyed in such a way that normal populations answer 95% of the questions in the keyed direction. Although the scale was not designed to measure personality variables, some personality indicators can tentatively be derived from this scale. This is based mainly on the observation that the content of the items reflects the following areas: good socialization, conformity, optimism, denial of neurotic characteristics, and conventionality of behavior and attitudes. Gough (1996) points out that this is

comparable to the "popular" response on the Rorschach in that it reflects the degree to which examinees see their surroundings in ways that are similar to others.

High Cm *(*T = *60 or More)*

High scores suggest that the examinee adheres to highly conventional attitudes and is overly socialized, tending to see his or her world in a stereotyped manner. These individuals do not see themselves as particularly unique or special and are conscientious and serious. They are most frequently described as clear-thinking, planful, practical, and tactful.

Low Cm *(*T = *30 or Less)*

A low score sometimes suggests that persons have chaotic, conflict-ridden family backgrounds. Their attitudes toward the world are typically unusual and idiosyncratic. They might also be generally upset, poorly motivated, self-defeating, frail, and lack a sense of meaning in life. The most frequent adjectives used to describe them are reckless, distractible, unconventional, moody, and confused.

However, scores in this range primarily increase the likelihood that the test is of questionable validity, and scores below 20 almost always confirm that the profile is invalid. The ideal score for detecting a "fake bad" profile is $T = 29$ for males and $T = 24$ for females (see Determine Profile Validity section).

13. Sense of Well Being (*Wb*)

The Well Being scale was originally developed to help recognize profiles in which the person was "faking bad." Thus, it was initially referred to as the Dissimulation (*Ds*) scale, and "fake bad" profiles can usually be detected because they are significantly lower than even valid profiles for psychiatric patients. In contrast, persons who score high do not have a need to emphasize psychological or physical complaints. In fact, high scorers play down their worries and, rather, emphasize that they are enterprising, energetic, and experience a sense of security. They are also likely to have effective interpersonal relations, a high level of mental health, and a sense of psychological and physical well being. Low scorers usually have diminished health and experience difficulty meeting the daily demands of their environment. In general, the *Wb* scale has come to represent a rough estimate of a person's level of adjustment and degree of psychological distress. However, it is more of a "state" scale than the others and is, therefore, somewhat changeable, depending on an individual's mood fluctuations.

The *Wb* scale has a low degree of item overlap with other scales because most of the questions were designed for exclusive use with this scale. The item content usually reflects a denial of various physical and psychological complaints. The second major content area reflects the extent to which a person is self-sufficient and independent.

High Wb *(*T = *55 or More)*

Generally, high scorers on *Wb* have relaxed and satisfying interpersonal relationships, are able to trust others, and come from family backgrounds that were stable and supportive. They are dependable, responsible, and value intellectual interests. Usually,

they are happily married and are stable, optimistic, and self-confident. The most frequent adjectives used to describe them are clear-thinking and capable.

Moderately Low Wb *(T = 35 to 50)*

Although persons scoring in this range generally feel that life is not going well, they continue to meet this perceived adversity with a sense of apathy and listlessness. They are often passive, awkward, and defensive.

Low Wb *(T = 35 or Less)*

With a further decrease in *Wb,* there is a corresponding exaggeration of the trends just discussed. These individuals are usually highly alienated and dissatisfied, and experience a significant level of maladjustment. Characteristically, they are extremely distrustful in interpersonal relationships, with a tendency to dwell on real or imagined wrongs. Such people are seen by others as pessimistic, tense, restless, and moody. They feel their life lacks a sense of meaning and might cope by becoming absorbed in fantasy and daydreams. Individuals who use the test situation as a forum for complaining and attempt to exaggerate their difficulties often score in this range. The most frequent adjectives used to describe these persons are confused, bitter, and nagging.

The interpretation of extremely low *Wb* scores requires two considerations. First, the scale lowering may in part reflect a downward but temporary mood shift of a person who is only somewhat maladjusted or even normal most of the time. More important, an extremely low score suggests an invalid profile in which the examinee is faking bad (see Determine Profile Validity section).

14. Tolerance (*To*)

The Tolerance scale was designed to measure the degree to which persons are socially intolerant versus the extent to which they are accepting, permissive, and nonjudgmental in their social beliefs and attitudes. The content of most of the items focuses on openness and flexibility versus rigidity and dogmatism. Other content areas relate to an interest in intellectual and aesthetic activities, level of trust, and a lack of hostility or resentment toward others. A person scoring high on Tolerance is also indicating that he or she is not alienated, does not feel isolated, rarely feels anxious, and is relatively poised and self-assured. There is a large variety of questions on this scale, but there is also a general lack of adequate validity studies. In fact, Tolerance is one of the poorer scales on the CPI, and its validity has even been questioned. Thus, interpretations based on this scale should be made cautiously and tentatively.

High To *(T = 60 or More)*

High scorers are likely to be intelligent, have a wide range of interests, and be socially tolerant. They are also able to trust others and may have a high degree of confidence and social poise. Furthermore, they are nonjudgmental, can easily accept divergent beliefs and values, and are forgiving, generous, and pleasant. They typically have a wide vocabulary and varied interests. They are concerned with philosophical issues and can effectively understand and explain the core of many problems. They are likeable and make a good impression because they are tolerant, permissive, and benevolent.

Extremely high scorers might be overly trusting to the extent that they are naive and underestimate potential difficulties. They might also be so worried about potential confrontations that they become overly adaptable and will fill any role to keep a situation peaceful. The most frequent adjectives used to describe them are fair-minded, insightful, clear-thinking, and interests wide.

Moderate To *(*T = *45 to 60)*

Moderate scorers are likewise somewhat nonjudgmental and open to the beliefs of others. They usually have a wide range of interest and are informal and independent.

Low To *(*T = *40 or Less)*

Persons scoring in this range are likely to be judgmental and nonaccepting of the beliefs and values of others. This judgmental attitude tends to generalize into other areas of their lives so that, overall, they seem cold, smug, and stern. They are authoritarian and center their lives on a fixed and dogmatic set of beliefs. Furthermore, they are mannerly, fearful, arrogant, and sarcastic. If criticized, they usually become extremely defensive, bitter, and rejecting. They are more likely to judge than to understand others. It is important for these individuals to exert power in relationships, and they may do so by becoming critical, outspoken, and holding unrealistic expectations. Internally, they often feel moody, distrustful, cynical, and dissatisfied. The most frequent adjectives used to describe them are prejudiced, interests narrow, and suspicious.

15. Achievement via Conformance (*Ac*)

The *Ac* scale involves not only an orientation toward achievement, but also a need for structure and organization as a means of channeling that achievement. This scale specifically relates to settings in which conformity is an asset and reflects the degree to which persons prefer to have their criteria of performance clearly specified by some outside source. The content of the items relates to how effectively they can perform in an academic setting and how high their relative levels of energy and efficiency are. High scorers also see themselves as being productive workers. Additional content areas relate to the extent to which the examinee is even-tempered, accepts the rules of socially-approved standards of behavior, and dislikes frivolous, unconventional behavior. Persons scoring low dislike externally imposed guidelines, are rebellious and disorganized (Gough, 1996), and are likely to experience career indecision (F. L. Newman et al., 1999).

The *Ac* scale has been one of the more thoroughly researched scales on the CPI, primarily because of its practical relevance for academic personnel. In a review of the literature, Megargee (1972) reports that it has good criterion validity and has been found to correlate significantly (.36 to .44) with GPA and general achievement in high school settings. The correlations are highest for high school performance and somewhat lower for college settings.

High Ac *(*T = *60 or More)*

Persons scoring above 60 are typically persistent and industrious, especially when conforming to some external standard. They strongly prefer specificity and structure,

and may even have a difficult time when structure is lacking, especially if a high *Ac* is accompanied by a low *Ai.* Such persons are usually responsible, capable, and ambitious, but they express these behaviors in a conservative, reserved, and obliging manner. Furthermore, they place a high degree of value on intellectual effort. They are most comfortable when working in highly organized settings, where they excel when given specific, well-defined criteria for performance. The most frequent adjectives used to describe them are responsible, organized, ambitious, persevering, efficient, and conscientious.

Moderately High Ac *(*T = *50 to 60)*

Moderate scorers may question the need for structure and organization. Although they may prefer not to have structure, they can adequately function in a structured situation when required to do so. They are usually stable, optimistic, dependable, and responsible.

Low Ac *(*T = *35 or Less)*

Persons in this range are rejecting of authority and regulations. This rebellion may result in achievements far below their potential because their energy is directed more toward rejecting external organization and rules rather than working within the limits imposed on them. Such persons are often characterized as intellectual rebels, especially if their Achievement via Independence (*Ai*) scale is relatively high. When external demands for performance are placed on them, they may become disorganized and nonproductive. They have difficulty committing themselves to organizations or people and, as a result, experience considerable caregiver indecision. The most frequent adjectives used to describe them are lazy, impulsive, reckless, rebellious, distractible, and mischievous.

16. Achievement via Independence (*Ai*)

Whereas *Ac* can be used to predict achievement in high school, *Ai* was designed to predict achievement in a college environment where independent initiative is more crucial. Persons who are high in *Ai* succeed in settings that require creativity, self-actualization, and independence of thought. Gough (1968) has clarified this distinction by describing Achievement via Conformance (*Ac*) as "form enhancing" whereas *Ai* is "form creating." *Ai* correlates significantly with college students' GPA (Gough & Lanning, 1986), yet there is only a low correlation with intelligence. Thus, students who have elevated *Ai* scales and who also achieve a high GPA do so mainly on the basis of a high need for achievement and only secondarily on the basis of intelligence. They are able to tolerate a high level of ambiguity and usually reject authoritarian or overly stringent regulations. In some cases, high *Ai* scores can predict achievement in situations in which originality and independence are rewarded. Persons with high scores are unwilling to accept conventional advice unquestioningly but rather prefer to think for themselves. Also, some questions relate to the degree to which individuals appreciate activities involving the intellect. Other content areas attempt to assess their degree of adjustment and the extent to which they are concerned with the deeper aspects of interpersonal relationships.

High Ai *(T = 60 or More)*

Such persons prefer to work without rules and structures, and usually feel restricted in a highly organized environment. They value creativity and originality and are self-motivated and rejecting of conventional standards of productivity. Their ability to produce and function is significantly impaired if a great deal of structure is required. They produce best and are most efficient when left to regulate their own behavior. Externally, they are verbally fluent, self-reliant, and make a good impression. They have a wide range of interests, high aspirations, and are concerned with philosophical interests. The adjectives used most frequently to describe them are intelligent, clear-thinking, logical, foresighted, insightful, and interests wide.

Moderate Ai *(T = 40 to 50)*

Persons scoring in this range are able to achieve based on their own self-direction but feel somewhat insecure when doing things completely on their own. Thus, they can work either with or without structure, but prefer a moderate degree of external organization. At times, they can be creative; but when they come to conclusions on their own, they still need external verification to feel comfortable.

Low Ai *(T = 35 or Less)*

Low scorers have difficulty trusting their own abilities, and this characteristic becomes more exaggerated as the scale score becomes lower. They require external definition to establish their self-concept and need others to specify their proper course of action. Because of this uncertainty and dependence on outside structure, these individuals are moderately anxious, depressed, and self-doubting. They are not intellectually inclined and tend to feel out of place in the world of abstract thinking. The adjectives most frequently used to describe them are confused and interests narrow.

17. Intellectual Efficiency (*Ie*)

The *Ie* scale was originally called a "nonintellectual intelligence test" and was designed to measure personality traits that coincided with a high level of intellectual ability. High scorers on *Ie* tend to be competent, clear-thinking, and to make efficient use of the potential they possess. Thus, it is less an intelligence test than it is a measure of the degree to which persons make efficient use of the intelligence they do possess. There is a moderate amount of item overlap with Sociability (*Sy*), Achievement via Independence (*Ai*), and Social Presence (*Sp*). One important content area of the items relates to the degree to which a person enjoys and is interested in wide-ranging intellectual activities. Also, a number of questions relate to self-confidence and assurance. Other questions relate to good physiological functioning, positive relationships with others, and an absence of irritability and suspiciousness.

A number of representative and noteworthy validity studies have been performed on *Ie*. It is positively correlated with measures of intelligence (Megaree, 1972), and members of MENSA scored significantly higher on *Ie* than the national norms (Southern & Plant, 1968). The scale has also been able to successfully discriminate high school dropouts from students who later graduated (Gough, 1966). The autobiographies of

high scorers reveal that they see themselves as well organized, efficient, and committed to pursuing intellectual and cultural activities (A. Hill, 1967).

High Ie *(*T = *60 or More)*

Persons scoring high on *Ie* have a wide range of interests with an excellent ability to use their resources. They are capable and confident, with good planning abilities, and are independent, informal, and clear-thinking. Their vocabulary is wide, and they are verbally fluent, perceptive, and effectively understand subtle nuances of behavior. They value intellectual activities and have high levels of aspiration. The most frequent adjectives used to describe them are intelligent, clear-thinking, alert, interests wide, and (having) initiative.

Moderate Ie *(*T = *40 to 60)*

Moderate scorers may still be highly competent, but they are also likely to have some self-doubts regarding their intellectual capabilities.

Low Ie *(*T = *40 or Less)*

Persons in this range may be insecure about their intellectual abilities, and are likely to experience enough self-doubt to create a mild degree of depression and anxiety. They typically appear awkward, shallow, and suggestible. They might give up easily and feel uncomfortable with uncertainties. As an alternative interpretation, low scorers may merely be uninterested in intellectual activities, which is also likely to be reflected in their choice of occupation. The latter interpretation would not imply the presence of self-doubt and insecurity suggested in the former, but rather merely a lack of interest. The most frequent adjectives used to describe individuals scoring low on *Ie* are confused, nervous, and interests narrow.

18. Psychological-Mindedness (*Py*)

The original intent of the *Py* scale was to identify persons who possess insight into the behavior of others in that they can accurately perceive the inner needs and motivations of others. This scale focuses on the ability to figure other people out and does not necessarily indicate people who are empathic and nurturing. To assess the degree of empathy of individuals, it was necessary to consult additional scales, such as Empathy (*Em*), Sociability (*Sy*), and Well Being (*Wb*). However, as further research was done on *Py,* it became clear that it was more an indicator of persons interested in pursuing psychology from an academic perspective. In fact, Megargee (1972) concludes his literature review by stating that the *Py* scale has limited usefulness as an indicator of a person's ability to accurately perceive the inner needs and motivations of others. The content of the items relates to a person's ability to concentrate, his or her effectiveness in dealing with ambiguity, and his or her degree of enjoyment in his or her occupation. Other content areas deal with an ability to stick with long-term goals and an acceptance of unconventional opinions.

High Py *(*T = *65 or More)*

High scorers are interested in academic pursuits, especially in the area of research. They can be highly original and creative in their approaches to abstract problems. They place a high level of importance on obtaining recognition for their efforts, and they

demonstrate perseverance, the ability to concentrate for long periods of time, and a high degree of satisfaction from their chosen profession. Other people often see them as independent, individualistic, preoccupied, and reserved. They are excellent in dealing with abstract situations but generally avoid concrete problem-solving situations. Extremely high scorers may be seen as distant, aloof, and detached. The most frequent adjectives used to describe them are logical, thorough, clear-thinking, foresighted, and interests wide.

Low Py *(*T = *35 or Less)*

Persons who score low on *Py* are generally not inclined toward research or scholarly activities. However, they are likely to be sociable, talkative, unassuming, and conventional. They usually accept the behavior and motivation of others at face value and are more comfortable with concrete situations. The most frequent adjectives used to describe them are simple and interests narrow.

19. Flexibility (*Fx*)

The *Fx* scale was designed to assess the degree to which an individual is flexible, adaptable, and changeable in his or her thinking, behavior, and temperament. It was originally based on questions relating to rigidity; but as the scale construction evolved, the scoring was reversed and the name changed from the Rigidity scale to the Flexibility scale. Other content areas relate to an ability to tolerate ambiguity, uncertainty, and impulsiveness, and to a nonjudgmental, tolerant attitude toward moral and ethical formulas of right and wrong.

The validity studies in part agree with the intent of the scale as they do support the hypothesis that low-scoring individuals are somewhat rigid. However, there is little evidence to indicate that extremely high scores reflect a high degree of flexibility (Megargee, 1972). Gough (1975) suggests that scores in the higher ranges are curvilinear—a moderately high score suggests that the person is relatively flexible, but with increasing elevation, a person becomes progressively more unstable and unpredictable. Megargee states that, given the weak evidence for the validity of this scale, especially for high *Fx,* it is one of the least valid scales on the CPI. Thus, any interpretations derived from it should be made with caution.

High Fx *(*T = *65 or More)*

Persons having extremely high scores may feel rootless and are often emotionally unstable. Everything in their lives is open to question, including their sense of values and moral beliefs. Thus, it is difficult for them to internalize clear-cut standards. They can easily approach situations from a number of varying perspectives. This allows them to consider many alternatives but may create a disadvantage in that they have difficulty developing a clearly defined direction. Extremely high scorers might be volatile, distractible, restless, and poorly organized. The most frequently used adjectives to describe high scorers are logical, thorough, clear-thinking, foresighted, and interests wide.

Moderate Fx *(*T = *50 to 65)*

Moderate scorers are open to considering and experiencing alternative perspectives. They are nonjudgmental, intellectually flexible, original, and able to develop innovative

ideas. They might also be independent, self-confident, optimistic, and value intellectual activities.

Moderately Low Fx *(*T = *35 to 50)*

Persons scoring in this range prefer structure and like to have things clearly defined and specified. Although they can handle a certain degree of uncertainty, it usually creates discomfort. They are usually cautious and practical and can be described as relatively rigid.

Low Fx *(*T = *35 or Less)*

Low scorers generally dislike new ideas and experiences and are continually seeking security. They have a strong need to control their thoughts and, generally, have a difficult time changing their decisions. They are usually rigid, stubborn, and defensive. Often, they have strong religious beliefs and are moralistic and conservative. Others perceive them as conscientious, serious, literal-minded, and overcontrolled. The most frequently used adjectives to describe them are organized, efficient, rigid, conservative, interests narrow, conventional, and prejudiced.

20. Femininity/Masculinity (*F/M*)

The *F/M* scale was developed to assess the degree to which examinees were psychologically feminine or masculine, regardless of their actual sex. Its original intent was to detect significant conflicts over sexual identity, but this aspect of the scale has become progressively less emphasized. The scale is currently used to assess the extent to which individuals endorse beliefs, values, and occupations that are traditionally held either by males or by females. The intent of some items is fairly obvious whereas the intent of other items is more subtle. Many items relate to traditional masculine or feminine roles. Additional content areas refer to a person's degree of restraint and impulsiveness, as well as the extent to which he or she is emotional during interpersonal relationships. McCrae et al. (1993) found that this was the only scale to correlate strongly (.45) with the core personality factor of Agreeableness. The items also reflect the degree to which a person is interested in politics, current affairs, and achievement. This scale has been well researched, and studies indicate it has a fairly high level of validity.

High F/M *(*T = *70 or More)*

For males, scores within this range suggest the possibility of difficulties related to sexual identity. These males might also be highly introspective and have philosophical and aesthetic interests. Their interests might be wide ranging, and their thought patterns unconventional. The most frequent adjectives used to describe them are nervous, worrying, weak, self-pitying, reflective, and sensitive.

Females with extremely high scores might be highly affiliative, dependent, submissive, and require continual reassurance. They might also be tolerant, permissive, giving, and oversensitive. Both male and female high scorers might use bodily symptoms to express anxiety and tension. The adjectives used most frequently to describe high-scoring females are warm, sympathetic, sentimental, and dependent.

Moderately High F/M *(T = 60 to 70 or More)*

Both males and females scoring within this range have significant needs for affiliation and dependency. They usually have a difficult time dealing with a high degree of autonomy and feel uncomfortable when independent action is required of them. They are both highly sensitive and quite concerned with not hurting others.

Moderate F/M *(T = 40 to 50)*

Persons scoring within this range can deal effectively with autonomy and have an average need for dependency and affiliation. They are generally practical and self-sufficient but not to an exaggerated extent.

Moderately Low F/M *(T = 40 or Less)*

Such persons are typically task oriented, practical, and emotionally self-sufficient with few dependency needs. They are often perceived as masculine, robust, tough, and even coarse.

Low-scoring males generally fit the masculine stereotype in that they are described as masculine, emotionally independent, tough minded, self-sufficient, and self-centered. They often have a clear, stable, internally consistent personality and adhere to conservative values. The adjectives most frequently used to describe low-scoring males are confident, independent, aggressive, and ambitious.

Females who score low are likewise self-reliant, confident, independent, and deliberate. In addition, they might also be critical, distrustful, cynical, and outspoken. They tend to be motivated by power and have high aspirations for themselves. The adjectives most frequently used to describe moderately low-scoring females are strong, tough, and independent.

Low F/M *(T = 30 or Less)*

An *F/M* score this low suggests an exaggeration of these trends and, in females, the likelihood of difficulties related to sexual identity.

SPECIAL PURPOSE SCALES

Several special purpose scales have been included in the Consulting Psychologists Press computer scoring and interpretive report and are summarized next.

Managerial Potential (*Mp*)

This scale identifies people who both seek and have talents in supervising and monitoring others' behavior. Such persons can think clearly, have effective interpersonal behavior, are confident, and are good at setting and following through with goals. At the same time, they do not exploit others and are not self-centered. High scorers (T = 60 or more) are trustworthy, efficient, productive, extroverted, well organized, mature, realistic, offer advice, value intellectual activities, make long-term plans, and have high aspirations. In contrast, persons with low scores (T = 40 or less) avoid making decisions, give up easily, are easily offended, apathetic, lack confidence, rebellious, dissatisfied, defensive, immature, become anxious with change.

Work Orientation (*Wo*)

Wo was designed to identify people who have a strong work ethic. Even in routine types of work, they are still likely to put forth their best efforts. Persons with high scores ($T = 60$ or more) are typically described as responsible, dependable, hard working, self-disciplined, moderate, reasonable, and require little praise or commendation as their attitude toward work is based more on internal values than external reinforcement. Low scorers ($T = 40$ or less) are likely to be distrustful, careless, self-centered, restless, and temperamental.

Creative Temperament (*CT*)

This scale identifies unconventional, artistic people who are likely to perceive the world in unusual, creative ways. High scorers ($T = 60$ or more) are usually described as imaginative, having a wide range of interests, like variety, react strongly to aesthetic material, have progressive social attitudes, and tend to be somewhat rebellious. Persons with low scores ($T = 40$ or less) are typically conventional, overcontrolled, reserved, rigid, have narrow interests, and prefer the status quo.

Leadership Potential (*Lp*)

This scale was designed to identify person who are effective at and feel comfortable in leadership positions. If persons score high ($T = 60$ or more), they are likely to be enterprising, confident, energetic, alert, ambitious, optimistic, resilient, resourceful, and can elicit cooperation from others. They have good initiative and are unlikely to be affected by criticism and pressure from others. In contrast, low scorers ($T = 40$ or less) can be described as temperamental, pessimistic, ill at ease, have feelings of inadequacy, worry excessively, have low self-confidence, and give up easily.

Amicability (*Ami*)

Ami identifies people who are friendly and considerate of others. Persons who score high ($T = 60$ or more) are ethical, consistent, cooperative, dependable, responsible, optimistic, warm, compassionate, and can easily establish close relationships with others. Some people might perceive them as submissive in that they are mild-mannered, self-controlled, and patient. Low scorers ($T = 40$ or less) are likely to be nonconforming, headstrong, self-centered, dramatizing, opportunistic, uncooperative, self-indulgent, and impatient. They can often be bitter, complaining, argumentative, manipulative, and irritable, particularly if demands are made on them.

Law Enforcement Orientation (*Leo*)

The intention and design of this scale was to identify people who both view societal rules and law enforcement favorably, and would also function well working in the law enforcement area. High scores ($T = 60$ or above) indicate someone who is optimistic, stable, ambitious, conscientious, and has leadership abilities. High scorers are also likely to be direct, honest, create a good impression, and have good interpersonal skills.

They are also conservative, conventional, moralistic, and hard-working. Persons with low scores ($T = 40$ or less) are described as nonconforming, cynical, distrustful, introspective, changeable, anxious, complicated, irritable, and may be self-defeating.

Tough-Mindedness (*Tm*)

The *Tm* scale measures where a person is on the continuum between tough-mindedness (frank, hard-hearted, unemotional) and tender-mindedness (trusting, reflective, soft-hearted). Persons scoring high ($T = 60$ or above) are likely to be independent, realistic, pragmatic, capable, determined, thorough, and confident. They are also hard-working, organized, have good leadership abilities, and can mobilize their resources quickly and effectively. They take pride in being objective, rational, and unemotional. When making decisions, they rely on facts rather than emotions. In contrast, persons with low scores ($T = 40$ or less) are likely to be sensitive, submissive, anxious, and easily feel inferior and embarrassed. Their motivation may be low, and they are likely to be undependable and have a difficult time dealing with stress.

CONFIGURAL INTERPRETATION

The following material on configural interpretation summarizes most of the empirical research on different code types. Regression equations have been included and are summarized in Table 9.2. The material is organized according to different topics (leadership, achievement, etc.). In contrast, McAllister (1996) has provided a listing of 152 code types arranged according to different patterns of low and high scale scores. Readers wanting to interpret scale scores can refer to McAllister or, alternatively, use the following topic listings. They might also wish to make rational interpretations of patterns of scale scores by using the following sequence:

1. Note the high (generally above 60) and low (generally below 40) scale scores and read the individual descriptions that correspond with these scores.
2. The key phrases that correspond with these single scales can then be written down, and the descriptions can be strengthened, weakened, or altered according to their relative elevations or lowerings and their relationships with other scales.
3. The descriptions can then be combined to create a more integrated description of the person.

Intellectual Level

Megargee (1972) has reported that *To, Ac, Ai, Ie, Py,* and *Fx* are all related to an individual's intellectual level. Elevations ($T = 55$ or more) on all or most of these scales strongly indicate that the person has a high interest in intellectual activities and good overall intelligence. Consistently low scores on all or most of these scales reflect limited intellectual ability and are a strong indication that the person has a narrow range of interests. This narrowing of interests may be, in part, a response to an emotionally upsetting event either in the recent past or at a significant time during the person's earlier development.

The particular patterns of high and low scales can provide information on the specific expression of intelligence. For example, it might be noted whether individuals would be more likely to excel in structured (high *Ac*) or nonstructured (high *Ai*) environments (see next subsection on achievement). Similarly, an interpreter can note how tolerant, flexible, or efficient individuals might be.

Achievement

Predicting and Assessing High School Achievement

The CPI is generally effective at detecting bright high school achievers. These students typically have elevated scores on *Ie* and *Ai,* whereas underachievers are generally low on these scales. Bright achievers also have relatively high scores on *Re, So, To, Ac,* and *Py.* Persons who are high achievers but have average IQs have relatively high scores ($T = 55$ or more) on *Re* and *So* and, to a lesser extent, on *Wb, Ac,* and *Ie.*

A number of equations have been developed for use in predicting achievement of high school students (see Megargee, 1972). These equations are composed of weighted combinations of scales and, when computed, provide the best possible prediction of specific abilities. For predicting the achievement of both males and females with combined low, medium, and high IQs, the following equation is recommended:

1. Achievement = 20.116 + .317 *Re* + .192 *So* − .309 *Gi* + .227 *Ac* + .280 *Ai* + .244 *Ie*

This equation correlates from .53 to .56 with overall high school GPA (Gough, 1964). If a student's IQ scores are available, the following equation is recommended:

2. Achievement = .786 + .195 *Re* + .244 *So* − .130 *Gi* + .19 *Ac* + .179 *Ai* + .279 IQ

Because *Ie* is a relatively inefficient measure of IQ, it has been excluded in this equation, and, instead, the exact IQ is from intelligence testing. This equation raises the correlation with overall GPA to .68, which is significantly better than the typical .60 (or lower) correlation found when using only IQ scores.

To evaluate whether students will drop out of high school or graduate and continue to college, social factors as measured by the CPI are at least as important as a student's intellectual ability. The primary scales used to predict high school graduation are *Re, Ac,* and, to a lesser extent, *Wb, To,* and *Ie,* all of which are usually significantly higher for students who graduate from high school than for students who are high school dropouts (Gough, 1964). High school students who later go to college score significantly higher on *Re, Ac,* and *Ie* (Gough, 1968). The following formula correlates at a level of .52 with later college attendance for high school students (Gough, 1968):

3. College Attendance = 17.822 + .333 *Do* + .539 *Cs* − .189 *Gi* + .740 *Ac*

Predicting and Assessing College Achievement

Several studies have been conducted on the relative importance of single-scale and combinations of scale scores in assessing college achievement. Significant correlations

have been found among *Re, So, Ai,* and overall GPA (Hase & Goldberg, 1967). Further studies (Flaherty & Reutzel, 1965; Griffin & Flaherty, 1964) likewise stress the importance of *Re, So,* and *Ai* but also include *Ie* and *Cs,* and, in female samples, *Do* was significantly correlated with GPA as well (Flaherty & Reutzel, 1965). These scales are somewhat similar to those used to predict achievement in high school students, except that *Ai* becomes more significant for college populations and *Ac* decreases in relative importance. In addition, the likelihood of later upward social mobility is correlated with *Cs* and college GPA.

Although positive correlations were found among these single scales, the magnitude of these correlations was not extremely high, with the highest correlation reaching only .36 for males on *Ai.* Most other significant correlations ranged between .20 and .26. However, weighted combinations of scores produced higher correlations ranging from .35 to .54, depending on the type of population being assessed. Gough (1964) has found a .41 correlation between the following formula and grades for both males and females in introductory psychology classes:

4. Achievement (Introduction to Psychology) = 35.958 − .294 *Sy* − .180 *Sp* + .185 *Re*
− .189 *Sc* − .152 *Gi* − .210 *Cm*
+ .275 *Ac* + .523 *Ai* + .241 *Ie*
+ .657 *Py*

Weighted combinations of scales in combination with SAT scores for males and females who were National Merit scholars were found to have a .32 and .23 correlation with college GPA, respectively:

5. Male GPA = .16 SAT (Math) + .11 *So* − .19 *Sp* + .17 *Fe*

6. Female GPA = .25 SAT (Verbal) − .14 *Sp* + .06 *Re* + .20 *Ac* + .08 *Fe*

Although the correlations derived from these formulas are somewhat low, they are an improvement on the use of SAT scores alone for this group.

Using a more general sample of college students' CPI scores to predict academic performance, Gough and Lanning (1986) found that *Ai, Ie,* and *Py* correlated at the levels of .28, .25, and .23, respectively. Multiple regression analysis produced the following equation, which had a correlation with later course grades of .38 for males and .36 for females:

7. GPA = 30.60 − .26 *Wb* + .35 *Re* − .19 *Gi* + .39 *Ai* + .22 *Ie* + .36 *Py*

Although the correlation was modest, it predicted academic performance somewhat better than using SAT-V (.31 for males, .38 for females) and SAT-M (.30 for males, .24 for females).

These rather modest correlations indicate that it is more difficult to predict performance for college students than for those attending high school. This can be traced to the far greater number and complexity of variables involved in a college setting. Both

the selection of curricula and the student's motivation for attending college can result from a variety of situations. Furthermore, significant changes have been made in the curricula and admissions policies of colleges since the early equations (4, 5, and 6) were developed. Finally, a student's lifestyle can be extremely varied. For example, some students may be attempting to struggle through college with a part- or even full-time job, whereas others may be taking relatively few classes and be supported exclusively by their parents. All of these variables are beyond the scope of what can be measured by a test such as the CPI. The practical implication for clinicians predicting college GPA is to consider not only test scores but also as many of the other variables as possible.

Achievement in Vocational Training Programs

Student Teaching

Several studies have been performed to assess the effectiveness of teachers in student-teaching programs. Veldman and Kelly (1965) found that student teachers who were rated highly by their supervisors scored significantly higher on *Ac, Cs, Do, Gi,* and *Py* than those who were rated as less effective. R. Hill (1960) also emphasized the importance of *Ac* but did not find *Do* and *Py* to be important. A further study with a female population again stressed the importance of *Ac* but also included *Re* and *Ie* as significant factors (Gough, Durflinger, & Hill, 1968). Although these studies consistently emphasized the importance of *Ac,* none of the other individual scales were found to have either consistent or large correlations with teaching effectiveness. However, Gough et al. (1968) found a moderate correlation of .44 between CPI scores and teacher effectiveness by using the following equation based on weighted scales:

$$8.\ \text{Teaching Effectiveness} = 14.743 + .334\,So - .670\,Gi + .997\,Ac + .909\,Py - .446\,Fx$$

Using this equation, they were able to predict with 65% accuracy the performance of student teachers.

Medical School

Several scales have been found to correlate positively with overall medical school GPA, including *Sy* (.35), *To* (.34), and *Ie* (.40; Gough & Hall, 1964). An equation based on weighted combinations of scores was found to correlate at a magnitude of .43 with both faculty ratings of students and GPA (Gough & Hall, 1964):

$$9.\ \text{Medical Promise} = .794\,Sy + .602\,To + 1.114\,Cm - .696\,Cs$$

Dental School

Most studies using single-scale correlations with achievement in dental school have not produced significant correlations, although Kirk, Cumming, and Hackett (1963) did report a correlation of .28 between *Ac* and dental school GPA. However, Gough and

Kirk (1970) found a .38 correlation with GPA by using the following equation based on weighted combinations of scales:

10. Dental Performance = 29.938 − .110 *Sp* + .148 *Re* − .262 *Gi* + .727 *Ac* + .230 *Py*

Although this correlation is somewhat modest, it is higher than the Dental Aptitude Test's correlation of .29.

Seminary

Query (1966) performed a study on seminary students who were advised to discontinue and those who successfully completed the program. Although he did not develop any equations based on weighted scores, he did find that those who were unsuccessful tended to score higher on *Sy* and *Sa.*

Police and Military Training

Both *Ie* (Hogan, 1973) and *Do* (Hargrave, Hiatt, & Gaffney, 1986) have been found to be related to police effectiveness. Hogan found that *Ie* correlated .40 with ratings of effectiveness made by instructors during training, and it correlated .43 after one year in training when ratings were made by field commanders (Mills & Bohannon, 1980). Other noteworthy correlations with other scales were for *Ac* (.31), *Ai* (.33), and *Sy* (.45; Hogan, 1971). Hargrave et al. (1986) described the most effective deputies as sociable, outgoing, and gregarious, whereas effective traffic controllers were characterized by a high capacity for rewarding social interactions. The most effective persons in both these groups (deputies and traffic controllers) were relatively dominant (high *Do*), energetic (high *Ie*), competitive (high *Ac*), independent (high *Ai*), flexible (high *Fx*), and socially ascendant (high Class 1 scales; Hargrave et al., 1986). Mills and Bohannon (1980) somewhat similarly described effective officers who had been in the field a year or more as independent (high *Ai*), energetic (high *Ie*), and flexible (high *Fx*). Police officers who later developed disciplinary problems were found to have lower scores on *Re, Sc,* and *So* than those who did not (Sarchione, et al., 1998).

Pugh (1985) has pointed out that what determines successful police performance changes over time. During their training and first year of employment, the most effective officers were found to be those who were most able to obtain the trust of their coworkers and become an accepted member of their department. After two years, their ability to strive for improvement (high *Cs*) became the best predictor. In contrast, the best predictors after 4.5 years of employment were qualities that indicated a person was stable, socially skilled, and responsible (high *Wb, Re, So*).

A study by D. Collins (1967) rated drill sergeants in a training program on the following four criteria of success: academic grades, an assessment of leadership ability, final class standing, and a field test of combat skills. The only scale to correlate significantly was *Ie.* It is interesting to note that the scales stressing conformity (*Ac*) and dominance (*Do*) had no correlation. This is in contrast to the frequent stereotype of drill sergeants as authoritarian, rigid, conformist, and autocratic. It has also been found that women who were successful in Air Force basic training scored higher on all scales except *Sc, Cm, Py,* and *Fe* than those who were unsuccessful (Elliot, 1960). A different study found that successful students graduating from an Army language

training program scored significantly higher on *Ai* and *Ie,* but not *Ac,* than those who were unsuccessful (Datel, Hall, & Rufe, 1965).

Achievement through Conformance versus Independence

A comparison between *Ac* and *Ai* can provide useful information regarding an individual's typical style or preference toward working. This can have important implications for helping a person make a career choice or understanding existing job difficulties. If *Ai* is high ($T = 55$ or more) and significantly higher than *Ac* (10 or more), such persons usually place a high level of trust in their own judgments and conclusions and are likely to reject conventional formulas. Their acceptance of decisions or ideas depends more on inward verification rather than a respect for, or adherence to, external standards. When left on their own, they are highly motivated to achieve, but they may feel restricted if placed in a structured environment. If *Ai* is exceptionally high ($T = 65$ or more), they may spend much of their time rejecting authority. This trend would be further exaggerated with high scores on *Do* and low scores on *Sy.* The result might be an almost obsessional quality in their thinking, characterized by strong themes of rebelliousness. In general, a significantly higher *Ai* than *Ac* is an excellent profile for authors, researchers, and persons in positions of independent leadership.

If *Ac* is high ($T = 55$ or more) and is significantly greater than *Ai* (10 or more), the opposite trend would be apparent. These persons would strongly prefer specificity and external structure. They would be more effective and feel more comfortable when "second in command," such as in a middle-management business position. An overall and generally effective combination occurs with high but evenly balanced scores on *Ai* and *Ac.* This suggests these individuals have the necessary flexibility both to work in a structured environment and to do effective work independently. The following is a listing of the descriptions given to persons scoring with different high and low combinations of *Ac* and *Ai* (Gough, 1968):

Ac High

idealistic	mannerly	intelligent	logical
cautious	shy	rational	interests are wide
praising	conscientious	realistic	inventive
nervous	inhibited	independent	active
helpful	dull	reasonable	stable

Ai Low — Ai High

irresponsible	show-off	spunky	tolerant
careless	touchy	reckless	reliable
distrustful	undependable	unexcitable	courageous
disorderly	unstable	foresighted	distractible
indifferent	restless	frank	pleasure-seeking

Ac Low

Leadership and Managerial Style

The CPI has been extensively used with organizations to understand leadership and managerial style. Additional uses have been in team building, consulting with executives, personnel selection, individual development, and filling recently vacated positions with persons within the organization (succession planning). An excellent, practical, case-focused guide toward using the CPI in these areas has been developed by P. Meyer and Davis (1992). They emphasize that one of the crucial issues to understand is that a person's optimal performance in an organization depends not only on personality, but also the degree to which an individual's personality fits into the direction the organization is headed as well as the organizational climate and culture. Thus, any interpretation of the CPI needs to take these factors into consideration. For example, an organization that is quickly undergoing extensive change will benefit from persons who are flexible and adaptable as might be reflected in high scores on the Flexibility scale. In some organizations, a consensus-building style is most appropriate; this might be suggested by high scores on dominance, tolerance, empathy, and *F/M* but with a lower score on social presence. In a fast-paced, high-risk organization, this same style might be quite dysfunctional. Thus, any interpretation of CPI profiles needs to integrate test results with the specifics of the organization.

Relevant information on leadership and managerial style can often be found by carefully considering the CPI profile. One way of guiding the questions that can be asked regarding leadership is to consider the following managerial competencies outlined by P. Meyer and Davis (1992):

Leadership orientation: Drive for influence, method of working with others to achieve goals, negotiation skills, willingness to take charge, forceful versus low-key style, authoritative versus consensus-building style.

Problem solving: Level of decisiveness, method of analyzing problems, extent to which they use others' input, cautiousness versus impulsivity, likelihood of considering a wide number of alternatives or holding to one or two options and arguing for them, creativity, and independence.

Achievement motivation: The extent or drive to achieve as well as the manner in which the person is likely to achieve, need for approval, need for recognition, achievement through independent efforts (form creating) versus working with and under the direction of others (form enhancing).

Interpersonal skills: Social comfort, extent to which they like to interact, awareness of interpersonal dynamics, concern and support for others, willingness to help and support others, tact, diplomacy, lack of abrasiveness, political astuteness.

Administrative skills: A person's orientation toward planning; need for structure, organization, and planning; attention to details; monitoring and controlling their behavior; focus on short-term as well as long-term planning; degree to which they approach work in a systematic manner.

Adaptability: Ability to cope with stress, self-reliance, ability to work in a wide number of contexts, tolerance of ambiguity, personal mastery, optimism, ability to be self-directed, possession of a wide number of adaptive behaviors.

Thus, managers might vary on the extent to which they need to take control, carefully consider all options in solving a problem, achieve individually or through conforming to some outside structure, are comfortable with their coworkers, awareness of details, and flexibility. The different high and low points on their CPI results can help to understand and elaborate on these differences in managerial style.

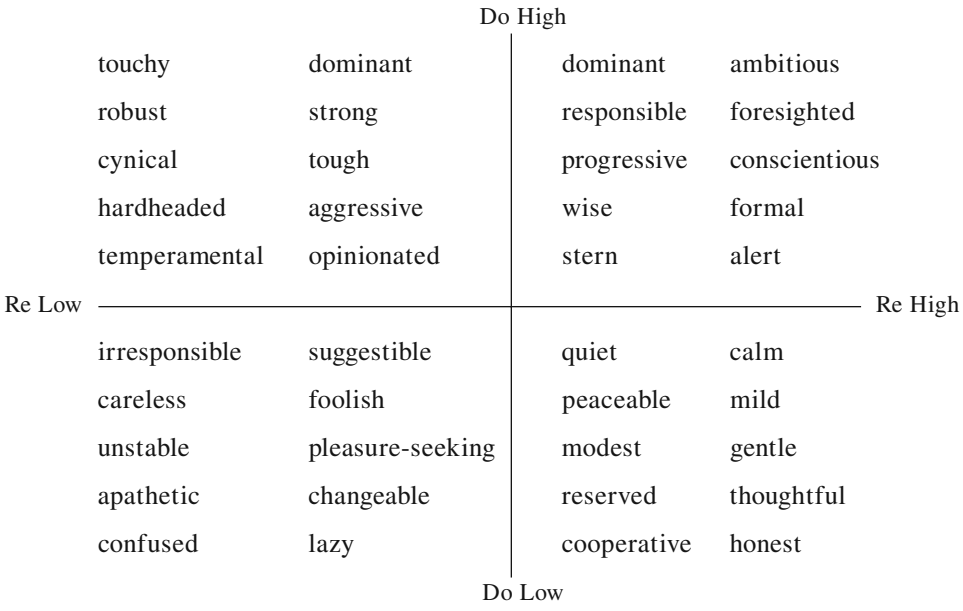
A general overview of a person's managerial and leadership potential can be noted by interpreting the Leadership and Managerial Potential special scales. This can be further refined by understanding that a person's comfort and wish to lead is largely related to his or her level of dominance. Accordingly, the *Do* scale has consistently proven to be accurate in differentiating leaders from nonleaders. In discussing leadership, it is helpful to describe the difference between an executive leader who has been appointed and a social leader who has been elected. For both types of leaders, the *Do* scale is high. However, for the executive leader, there is considerably more variability among the other scales; the style of expressing leadership is more dependent on the conditions the person is in, and the achievement scales are relatively more important than the other measurements (summarized in Megargee, 1972). This seems reasonable because the success of an executive leader is based more on his or her administrative and supervisory abilities than on his or her popularity. Social leadership is more likely to have a general elevation in Factor 2 scales as well as an elevated *Do*.

For example, if *Do, Cs,* and *Sp* are the high points, the leader is likely to be socially charismatic, persuasive, at the center of attention, and energetic (Heilbrun, Daniel, Goodstein, Stephenson, & Crites, 1962). If *Do* is high along with *Sa* and *Ac,* the person will have a high need for control, fear rejection, demand attention, dislike surprises, and emphasize clear structure (McAllister, 1996). If *Do* and *Ai* are the high points, these individuals are independent achievers who may also be highly creative self-initiators (McAllister, 1996). Further interactions with *Do* can likewise be developed by taking into consideration the specific meanings of additional corresponding high and low scale scores.

Using a combination of weighted scales derived from social leaders in a high school environment, Gough (1969) was able to obtain a modest correlation of .34 between social leadership and weighted CPI scales.

11. Leadership (Social) = 14.130 + .372 *Do* + .696 *Sa* + .345 *Wb* − .133 *Gi* + .274 *Ai*

Gough (1968) studied the relationship between *Do* and *Re* and found that the meaning of *Do* is altered by the relative elevation of *Re*. If *Do* and *Re* are both high, a leader is generally progressive, conscientious, and ambitious. In contrast, adjectives describing high *Do* persons with low *Re* indicate that they are dominant in a more aggressive, rigid, and destructive way. The following is a list of adjectives used to describe various combinations of *Do* and *Re:*



Executive Success

Success and effectiveness as an executive are frequently found in a profile in which $T = 60$ on *Do, Cs,* and *Sp;* $T = 40–50$ on *Sa, Re, So, Sc;* $T = 55$ or more on *Sy;* $T = 40$ or more on *Wb;* and $T = 50$ or less on *Gi* (in Webb, McNamara, & Rodgers, 1981). The most important variables are the indicated *T* scores on *Do, Cs, Sp, Sa, Re, So,* and *Sc.* This profile is common among business executives and managers. They are usually able to have others adapt to their plans, yet, at the same time, are flexible enough to adapt to the demands that are placed on them. Although they are generally excellent leaders, they may create a certain degree of family discord by attempting to be too demanding and autocratic in the home. If this combination of scores is present for a person under 25 years of age, it can suggest a naive sense of overconfidence in which the person cannot effectively assess his or her personal limitations. However, this profile is generally a good predictor of later success in leadership positions.

Leadership and Empathy

If an individual has elevations on both *Do* ($T = 65$ or more) and *Gi* ($T = 60$ or more), he or she is likely to not only possess excellent leadership abilities, but also to demonstrate a concern with, and empathy for, others (Heilbrun et al., 1962). If *Gi* is low in relationship to *Do,* the leadership style is usually more critical, domineering, egotistical, and autocratic, with a decreased concern for creating and maintaining harmonious interpersonal relationships in the group. A low score on both *Gi* and *Do* reflects a somewhat passive and withdrawn person who is socially inept and resentful, and whose passivity may be expressed in a shy seeking of approval from others.

Decision Making

The interaction between *Sa* and *Wb* reflects the degree to which the examinee turns to himself or herself for decision making or depends on others. If *Wb* is low and *Sa* is

moderate to high, these persons usually rely on their own self-evaluations and feel that others are inferior and cannot be trusted. This may result from their self-assurance and independence as reflected by a *Wb* that is only moderate to slightly low, or they may listen to only their own judgments because of a deep sense of alienation and distrust of others, as reflected in a markedly low *Wb* and high *Sa.*

If *Wb* is moderate to high and *Sa* is low, such individuals tend to believe that the judgment of others is superior to their own judgment, perhaps because they are still fairly accepting of themselves (only slightly low *Sa*) but think even more highly of others. Thus, they may have a high level of loyalty to people who are in superior positions, such as an employer or parent. Such persons may also have a poorly developed ability to accurately perceive the faults and limitations of others, and may have developed this loyalty in response to overprotective parents. A further possibility could be that they do not respect their own judgments and perceptions because they are lacking in their own resources.

With both *Sa* and *Wb* low, a person is likely to have significant doubts regarding himself or herself. There may be an excessive level of dependency, fearfulness regarding his or her own competence, and a corresponding resentment of his or her continual dependency on others.

Clinical Assessment

The CPI has generally not proven to be as effective in the assessment of psychopathology as it has in the educational and vocational areas. This can be traced to several reasons, but, primarily, it was not designed for clinical assessment and thus relatively little research has been conducted in this area. The organization and nature of the scales were not designed to differentiate among the various syndromes of pathology, nor do they provide information relating to a person's intrapsychic areas of functioning. Furthermore, devices such as the MMPI-2 and MCMI-III are clearly superior for the evaluation of pathology.

Despite these limitations, the CPI can make some general, as well as specific, contributions. Even though it does not distinguish between the different patterns of pathology, general maladjustment is usually indicated by lowered profiles (Higgins-Lee, 1990). The CPI has also been effectively used to detect and assess criminal and delinquent individuals, which involves a more interpersonal or, more accurately, an individual versus societal type of conflict. Furthermore, the CPI is a good adjunct to more clinically-oriented tests, because it can assess the relative strengths in an otherwise pathological individual and answer questions relating to the type of educational and vocational programs this person might benefit from.

General Maladjustment

An individual's level of maladjustment is indicated by generally lowered profiles (Higgins-Lee, 1990), which are often accompanied by an elevation on *Fe.* The scales found to best predict personal distress were low scores on *Ie, Mp, So,* and *Py* (Cook et al., 1996). A lowering of Factor 1 scales (especially *Do, Re, So,* and *Sc*) is often a good indicator of poor adjustment, and men with low *Ac* and *Ie* are especially likely to be maladjusted (Stewart, 1962).

Personality Disorder

Persons with scores below 25 on the *So* scale are likely to have diagnoses of personality disorders, particularly those that are related to dramatic, emotional, or erratic behaviors (borderline, antisocial, histrionic, narcissistic; Kadden et al., 1996; Standage, 1990; Standage, Smith, & Norman, 1988). However, if they score high on items related to "problem behaviors" (indicating denial of these problem behaviors), but low on items reflecting dysphoria and having had an unhappy childhood (thereby agreeing to these difficulties), they would be unlikely to have alcohol or drug problems but may have difficulties with depression. The likelihood that they would have personality disorders would then be intermediate (Standage et al., 1988).

Vulnerability to Stress

Persons with a "V" formation in which *So* is low ($T = 35$ or less), with *Re* and *Sc* significantly higher ($T = 40$ or more), are likely to be defensive and susceptible to the effects of situational stress (in Webb et al., 1981). They usually come from chaotic, stress-filled families in which there were episodes of irrational parental abuse. Thus, they have learned that the world is a dangerous place and have developed a precarious balance in which they feel constantly on guard. They keep their emotions carefully controlled, continually attempt to avoid conflict, and feel they need to be constantly prepared to diffuse potentially stress-filled interactions. Their conformity to their environment is based not on an expectation of achieving positive rewards, but more on fear and an avoidance of negative consequences. These people may have occasional explosive outbursts in which they have an almost dissociative loss of control. This explosiveness is especially likely if their spouses are manipulative, insensitive, and exploitive. As the discrepancy between *So* and *Sc* increases, these dynamics become more pronounced.

Depression

The Social Ascendency (Class 1) scales are generally lowered by depression, and a *T* score of 40 or less on *Sy, So, Wb,* and *Ie* is highly typical of depressed populations (Holliman & Montross, 1984). The scales that provided the best indicators of depression in males were *Sy, So,* and *Ie,* whereas for females the best discriminators were *Wb, So,* and *Ie* (Holliman & Montross, 1984). In most cases, the *Wb* scale is particularly important to notice, because a lowering on scales such as *Do, Cs, Sy, Sp,* and *Sa* might suggest merely a shy, unassertive, socially uninvolved person who is not necessarily depressed (McAllister, 1996). When the depression begins to lift and the person starts to have more optimism and a greater orientation to his or her environment, these scales generally increase. The mental and behavioral apathy often associated with depression can also be reflected by a lowering ($T = 40$ or less) in *Ac, Ai,* and *Ie.*

Psychosomatic Disorders

Although the CPI was not designed to diagnose psychosomatic disorders, it can assess certain personality characteristics that are consistent with individuals who are susceptible to this type of disturbance. Both male and female psychosomatics usually have lowered scores ($T = 40$ or less) on *Wb* and *Sc* and an elevation on *Cm* (Stewart, 1962). In addition, males often have a lowering on *Ie.* When the scores from male and female

psychosomatics are compared with persons having behavior disorders, psychosomatics have a relatively higher *So* and *Cm,* with females also having a higher *Re* (Stewart, 1962). All of these scores suggest that psychosomatic patients have a significantly higher level of superego control and socialization. This agrees with most formulations of psychosomatic disorders that emphasize the suppression and repression of hostility and antisocial behavior as important predisposing factors. A pattern of psychosomatic disorders is especially likely if *Wb* has a *T* score of 35 or less, accompanied by an *Fe* of 60 or more. This pattern is associated with headaches, gastrointestinal upsets, or functional skin conditions. Such persons are likely to have moderately high needs for dependency, which are not being fulfilled, but they also tend to feel distrustful and alienated in their relationships with others.

Defense Mechanisms

The two basic approaches to defense are either through repression or through sensitization. Whereas repressors attempt to avoid anxiety-arousing stimuli, sensitizers approach and attempt to control situations. Byrne, Golightly, and Sheffield (1965) found that high scorers on *Sy, Wb, Sc, To, Gi, Ac,* and *Ie* were more likely to use repression.

Certain types of assaultive offenders can usually be characterized as overcontrolled, but occasionally they drop all inhibitions and impulsively strike out (Megargee, 1964, 1965, 1966). These persons score high on the Overcontrolled Hostility scale (OH) of the MMPI and, also, have higher scores on *Sc* and *Gi* with a lowering on *Sa* (Megargee et al., 1967). This gives further support to the view that *Sc* and *Gi* are associated with the use of repressive defenses.

Juvenile Delinquency and Criminal Behavior

The assessment of antisocial behavior with the CPI has been well researched with generally useful findings. Both delinquents and criminals tend to have lower overall subscale scores, particularly on *Re* and *So* (J. Collins & Bagozzi, 1999; Gough & Bradley, 1992; Laufer, Skoog, & Day, 1982). In addition, persons who scored in either the Gamma or Delta lifestyle categories (questioning of normative beliefs) were more likely to be delinquents than those in the Alpha or Beta categories (Gough & Bradley, 1992). However, level of self-realization or integration (score on Vector 3) is important to consider in that delinquents and criminals had low levels of self-realization. In contrast, persons with high levels of self-realization were found to have a low level of criminal or delinquent behavior regardless of which of the four lifestyles they were in (Gough & Bradley, 1992).

Scores on *Wb, To,* and *Ac* are also likely to be somewhat lower (Gough & Bradley, 1992). This pattern suggests that the social poise of delinquents is usually about the same as that of other persons their age; but, in most other respects, their behavior is definitely unconventional, and they usually do not channel these differences into creative or intellectual areas. Mizushima and DeVos (1967) have found significant differences on the CPI between solitary delinquents who have lower scores on *Ie* and *Fe* and more socially-oriented delinquents who have significantly higher scores on *Sy, Sp,* and *Sa.* They also found violent offenders to be higher on *Sp* and *Sa* but low on *Fe.* However, delinquents who committed extremely violent offenses were especially high on *Sc,* which supports Gough's theory that excessive overcontrol in certain individuals periodically breaks

down, leading to assaultive behavior (Megargee, 1966). In summarizing this data on delinquency, it is most important to consider lowerings in *Re* and *So*. Further information regarding the style of delinquency can be derived by the lowered *Ie* and *Fe* for solitary delinquents; higher *Sy, Sp,* and *Sa* for social delinquents; higher *Sp, Sa,* and low *Fe* for violent social delinquents; and outstandingly high *Sc* for extremely violent offenders who have periodic excessive losses of control.

The likelihood of successful parole for delinquents can, in part, be predicted in that more successful parolees have higher scores on *Sp* and *Sa;* and less successful parolees have lower scores on *So* and *Sc* (Gough, Wenk, & Rozynko, 1965). Gough and his colleagues have developed the following regression equation to differentiate successful from unsuccessful parolees:

$$12.\ \text{Parole Success} = 45.078 - .353\,Sp - .182\,Sa + .532\,So + .224\,Sc$$

Using this equation, Gough et al. (1965) were able to predict with 60% accuracy which of a population of California Youth Authority parolees would be successful and which parolees would later become recidivists.

Marital Violence among Males

Males who were found to be physically violent in their marriages generally had lower scores in ten of the CPI scales including *Re, So, Sc, To, Ac, Ai, Gi, Ie,* and *Py* (Barnett & Hamberger, 1992). Low scores on *Re, So,* and *Gi* were particularly good predictors and suggested that persons with these scales as their lowest scores had difficulties with impulsivity, problem solving, and intimacy.

Chemical Dependency

Kurtines, Hogan, and Weiss (1975) found that the possibility of potential or actual substance abuse, perhaps to the extent of actual addiction, is suggested by high *Sp* and *Sa* accompanied by low scores on *Re, So, Sc,* and *Wb.* In a college population, J. Goldstein (1974) somewhat similarly found that students who used drugs were more likely than nonusers to have elevations on Capacity for Status, Social Presence, Self-Acceptance, Psychological-Mindedness, and Flexibility. In contrast, nonusers scored higher on Well Being, Responsibility, Socialization, Self-Control, Tolerance, Communality, Achievement via Conformance, and Intellectual Efficiency. The scale with the greatest ability to differentiate the two groups was Socialization, with users having a mean of 41.2 and nonusers having a mean of 52.1.

Social Maturity

The concept of social maturity includes *So,* but is more extensive and also includes areas other than that assessed by the *So* scale alone. Specifically, the person who is considered to be socially mature is not merely directed by blind conformance, but also has a high level of ethical standards that can even vary from the values held by the majority of people. He or she may, at times, feel a need to resist social pressure. In addition, this person can accurately perceive the faults in a social system and attempt to deal with them in a mature way. Thus, the socially mature person is clearly different

from someone who is merely oversocialized or hypernormal. Gough (1966) developed the following multiple regression equation to assess social maturity using combined weighted scores:

$$13.\ \text{Social Maturity} = 25.701 + .408\,Re + .478\,So - .296\,Gi$$

A special Social Maturity scale has also been developed and is available as a portion of the CPI computer report (see Gough, 1996; McAllister, 1996).

RECOMMENDED READING

Gough, H. (1996). *California Psychological Inventory manual.* Palo Alto, CA: Consulting Psychologists Press.

McAllister, L. (1996). *A practical guide to CPI interpretation* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Megargee, E. I. (2002). *The California Psychological Inventory handbook* (2nd ed.). San Francisco: Jossey-Bass.

Meyer, P., & Davis, S. (1992). *The CPI applications guide: An essential tool for individual, group, and organizational development.* Palo Alto, CA: Consulting Psychologists Press.

Chapter 10

THE RORSCHACH

The Rorschach has traditionally been considered a projective test consisting of a set of 10 bilaterally symmetrical inkblots. Subjects are asked to tell the examiner what the inkblots remind them of. The overall goal of the technique is to assess the structure of personality, with particular emphasis on how individuals construct their experience (cognitive structuring) and the meanings assigned to their perceptual experiences (thematic imagery; Weiner, 1994). The interpretations on Rorschach data can provide information on variables such as motivations, response tendencies, cognitive operations, affectivity, and personal and interpersonal perceptions. Despite attacks from both in and outside the field of psychology, the Rorschach remains one of the most extensively used and thoroughly researched techniques (Archer & Newsom, 2000; Camara et al., 2000; Watkins et al., 1995). This is reflected in the fact that more than 200 books and 9,000 articles have been written about or using the Rorschach (Exner, 1997).

The central assumption of the Rorschach is that stimuli from the environment are organized by a person's specific needs, motives, and conflicts, and by certain perceptual "sets." This need for organization becomes more exaggerated, extensive, and conspicuous when subjects are confronted with ambiguous stimuli, such as inkblots. Thus, they must draw on their personal internal images, ideas, and relationships to create a response. This process requires that persons organize these perceptions as well as associate them with experiences and impressions. The central thesis on which Rorschach interpretation is based is this: The process by which persons organize their responses to the Rorschach is representative of how they confront other ambiguous situations requiring organization and judgment. Once the responses have been made and recorded, they are scored according to three general categories: (a) the *location,* or the area of the inkblot on which they focused; (b) *determinants,* or specific properties of the blot they used in making their responses (color, shape, and so on); and (c) the *content,* or general class of objects to which the response belongs (human, architecture, anatomy, etc.). The interpretation of the overall protocol is based on the relative number of responses that fall into each of these categories. Some systems also score for the extent to which subjects organize their responses (organizational activity), the types of verbalizations, and the meaningful associations related to the inkblots.

Although these scoring categories may appear straightforward, the specifics of scoring and interpreting the Rorschach are extremely complex. Furthermore, attempts to develop a precise, universally accepted coding system have not been entirely successful, which creates some confusion and ambiguity in approaching the Rorschach technique itself (see Weiner, 1994). Although the primary scoring systems have some agreed-on similarities, there are also significant differences in the elements of these

systems. These differences, in turn, reflect the complexity and ambiguity in the nature of the responses made to the cards. Thus, effective use of the Rorschach depends on a thorough knowledge of a scoring system, clinical experience, and adequate knowledge of personality and psychopathology.

The general purpose of this chapter is to provide an overview of administration, scoring, and interpretation using Exner's Comprehensive System. Exner's system was selected because it is the most ambitious and psychometrically sound Rorschach system to date. Furthermore, the most frequently used scorings and interpretations from other systems have been included and integrated into Exner's approach.

Scoring for the Comprehensive System is quite complex, and only a brief overview can be covered in this chapter. Clinicians who wish to use precise scoring tables and criteria, as well as more extensive elaborations on interpretation, are encouraged to consult Exner and his colleagues' original works (Exner, 2000, 2001, 2003; Exner & Weiner, 1995). This chapter cannot stand as a substitute for Exner's work. Its major intent is to familiarize persons with the Rorschach in general and, more specifically, with Exner's approach to interpretation. This emphasis on interpretation rather than the technicalities of scoring is consistent with a recent survey noting that the main area in which students experience insufficient training is the actual interpretation and integration of results (Hilsenroth & Handler, 1995). In addition to students who are learning the system, persons who are already familiar with Exner's system might wish to consult sections of this chapter to obtain summaries of different scoring categories and interpretive hypotheses. This might be most appropriate for practitioners who use the Rorschach only occasionally. Finally, persons who use other scoring systems may wish to consult the different interpretive hypotheses as an aid to interpretation. This is theoretically possible because Exner incorporated the major approaches from other systems into his Comprehensive System. However, many minor variations are likely to occur between the Comprehensive System and other systems; therefore, interpretations should be made with caution.

HISTORY AND DEVELOPMENT

Many inkblot-type tests and games had existed long before Rorschach published his original 10 cards in 1921. For example, da Vinci and Botticelli were interested in determining how a person's interpretations of ambiguous designs reflected his or her personality. This theme was later considered by Binet and Henri in 1895, and by Whipple in 1910. A popular parlor game named Blotto that developed in the late 1800s required players to make creative responses to inkblots. However, Rorschach developed the first extensive, empirically based system to score and interpret responses to a standardized set of cards. Unfortunately, Rorschach died at age 37, shortly after the publication of his major work, *Psychodiagnostik* (1921/1941). His work was continued to a limited extent by three of his colleagues—Emil Oberholzer, George Roeurer, and Walter Morgenthaler.

The main approach used by Rorschach and other early developers of inkblot techniques was to note the characteristic responses of different types of populations. Thus, the initial norms were developed to help differentiate among various clinical

and normal populations: schizophrenics, persons with intellectual disabilities (mentally retarded), normals, artists, scholars, and other specific subgroups with known characteristics. Rorschach primarily wanted to establish empirically based discriminations among different groups and was only minimally concerned with the symbolical interpretation of contents. Many of his original concepts and scoring categories have been continued within current systems of analysis. For example, he noted that depressed, sullen patients seemed to give the fewest responses. Persons giving a large number of very quick responses were likely to be similarly "scattered" in their perception and ideation to nontest situations. He also considered the importance of long latencies (so-called "shock" responses), and hypothesized that they were related to a sense of helplessness and emotional repression.

Had Rorschach lived longer, the history and development of his test might have been quite different. Without the continued guidance and research from the "founding father," the strands of the Rorschach technique were taken up by persons who had quite different backgrounds from Rorschach and from one another. By 1957, five Rorschach systems were in wide use, the most popular being those developed by Beck and Klopfer. These two approaches came to represent polarized schools of thought and were often in conflict.

S. J. Beck (1937) adhered closely to Rorschach's format for coding and scoring. He continually stressed the importance of establishing strong empirical relationships between Rorschach codes and outside criterion measures. Beck emphasized that the response to the Rorschach involved primarily a perceptual-cognitive process in which the respondents structure and organize their perceptions into meaningful responses. This perceptual-cognitive process was likely to reflect their responses to their world in general. For example, persons who broke down their perceptions of an inkblot into small details were likely to behave similarly for perceptions outside the testing situation.

In contrast, B. Klopfer (1937) was closely aligned to phenomenology and the theories of personality developed by Freud and Jung. As a result, he emphasized the symbolical and experiential nature of a respondent's Rorschach contents. Thus, Klopfer believed that Rorschach responses were fantasy products triggered by the stimulus of the inkblots. For example, persons who perceived threatening objects on the inkblots would reflect persons who perceive aspects of their world as similarly threatening. Although not as popular, additional systems developed by Piotrowski, Hertz, and Rapaport represented a middle ground between the two extremes taken by Beck and Klopfer.

With five distinct systems available, the Rorschach became not a unitary test, but five different tests. Exner (1969) provided a comparative analysis of these different systems and later concluded that "the notion of *the* Rorschach was more myth than reality" (Exner, 1986, p. 19). He pointed out that none of the five systems used the same verbal instructions and only two of the systems required identical seating arrangements. More importantly, each systematizer developed his or her own format for scoring, which resulted in many differences regarding interpretation, the components required to calculate quantitative formulas, the meanings associated with many of the variables, and the interpretive postulates.

The wide range of often competing approaches resulted in numerous detrimental practices. A survey of practitioners by Exner and Exner (1972) indicated that 22% of all respondents had abandoned scoring altogether and, instead, based their interpretations

on a subjective analysis of contents. Of those who did score, 75% used their own personalized integration of scores from a variety of systems. In addition, the vast majority did not follow any prescribed set of instructions for administration. With researchers using a variety of approaches, comparison of the results of different studies was difficult. Researchers in the early 1970s further reported difficulties in recruiting subjects, problems with experimenter bias that needed to be corrected by using multiple examiners, statistical complexities of data analysis, inadequate control groups, and insufficient normative data (Exner, 1993, 2003).

The general conclusion, based on these findings, was that the research on and the clinical use of the Rorschach were seriously flawed, in part because of the lack of clarity inherent in having five different systems. Surveys and analysis of research conducted in the early 1970s by Exner and his colleagues concluded that, although all five systems included some empirically sturdy elements, they also included elements that had no empirical basis or elements for which negative findings were predominant.

To correct the difficulties with both the research and clinical use of the Rorschach, Exner and his colleagues began the collection of a broad normative database and the development of an integrated system of scoring and interpretation. Their initial step was to establish clear guidelines for seating, verbal instructions, recording, and inquiry by the examiner regarding the examinee's responses. The best features for scoring and interpretation, based on both empirical validation and commonality across systems, were adapted from each of the five different systems. A scoring category was included in the new system only after it had achieved a minimum .85 level for interscorer reliability. The final product was first published in 1974 as *The Rorschach: A Comprehensive System* and has since been released in second (1986), third (1993), and fourth (2003) editions. A second volume relating to current research and interpretation has been released in two editions (Exner, 1978, 1991), and two editions of a volume on the assessment of children and adolescents have also been published (Exner & Weiner, 1982, 1995).

Normative data for the Comprehensive System has undergone continual revision. A major reason for these revisions has been to refine stratification. A further impetus was that in 1990 the Comprehensive System eliminated all protocols with fewer than 14 responses because these were likely to have resulted in invalid protocols. The normative base reported in Exner's 1993 (3rd) edition of the Comprehensive System was composed of 700 adult nonpatients and 1,390 nonpatient children and adolescents between the ages of 5 and 16. However, it was discovered in 1999 that more than 200 duplicate adult protocols had inadvertently been included. These were replaced with new nonpatient protocols that had not been used previously and were selected so as not to alter the demographic features of the sample. The resulting normative sample has been published in the most recent (4th) edition of the Comprehensive System (Exner, 2003) and is composed of 600 adult nonpatients. The sample is evenly divided into males and females (300 in each group) with 120 subjects from each of five representative geographic areas. The mean age is 31.73 (SD = 10.69, Median = 30, Mode = 22, range = 19 to 69) with an average of 13.43 years of education (range = 8 to 19). A new project is currently in progress to completely revise the norms. Exner (2003) has reported the data from the first 175 participants and notes that, in most cases, there have been few discrepancies between the data from the earlier sample of 600 adults (as refined and reported in 2003)

and the new partially complete norms of 175 participants. The child and adolescent sample reported in Exner, 2003, is the same as that included in 1993 (includes 1,390 nonpatients between the ages of 5 and 16).

Exner's integration of the different Rorschach approaches into his Comprehensive System has been successful in that most research studies over the past 20 years have used his system, and it has become by far the most frequently taught system in graduate training (Hilsenroth & Handler, 1995). His attention to empirical validation, combined with a large normative database, has served to increase its acceptance and status. Access to training and interpretive aids has been facilitated through numerous workshops, a scoring workbook (Exner, 2001), ongoing research publications, new editions of earlier volumes, and computer-assisted scoring and interpretation (Exner, 1984, 1986, 1993, 2003).

Debates regarding the psychometric adequacy of the Rorschach have created one of the greatest controversies in the history of psychology. From the beginning, the Rorschach was met with skepticism in the United States; yet, it developed a strong following. At one point, the Rorschach was the second most frequently used test, and, in the 1940s and 1950s, the name *Rorschach* was almost synonymous with clinical psychology. Despite this initial (and continuing) popularity, reviews have generally been quite critical. As early as 1954, Shaffer declared that the Rorschach could no longer be considered a promising instrument and, 11 years later, Dana (1965) somewhat prematurely concluded: "Indeed, we have come to the end of an era, preoccupation with the Rorschach as a test" (p. 495). A. R. Jensen (1965) was even more critical when he recommended that "the Rorschach be altogether abandoned in clinical practice, and that students in clinical psychology not be required to waste their time learning the technique" (p. 509). Most recently, Garb (1999) has called for a "moratorium" on its use until research has clarified which scoring categories are valid.

It should be noted that one of the early difficulties in establishing the psychometric properties of the Rorschach was in making meaningful comparisons across various studies. As Exner (1969, 1974, 1986, 1993, 2003) has repeatedly pointed out, there is not *a* Rorschach; rather, at least five different Rorschachs have been created around the five major systems. Reliability and validity studies performed on one system did not necessarily mean that the findings from these studies could be generalized to any of the other systems. However, reviewers often acted as if there were only one Rorschach. Furthermore, many studies were poorly conducted. They were characterized by inadequate controls for age, sex, race, IQ, and socioeconomic status. In addition, many studies had extremely wide variations in the training required for scorers, insufficient protection from experimenter bias, poor validation criteria, and inadequate statistical models. These difficulties were amply demonstrated when Exner (1986) and his associates found it necessary to discard 1,400 research studies of a total of 2,100 studies published before 1970.

More recently, the depth and sophistication of the criticisms have increased. This has resulted in extensive arguments and counterarguments, with each side citing numerous studies in favor of their positions. In the late 1990s, this culminated in most major assessment journals' publishing special series debating the relative merits of the Rorschach. Challenges were directed at nearly all aspects of the Rorschach,

including the adequacy of its norms, interscorer reliability, temporal stability, the accuracy of meta-analysis that had found support for the Rorschach, and its level of incremental validity. The central elements of these debates are integrated into the sections on reliability and validity and assets and limitations.

RELIABILITY AND VALIDITY

As noted previously, Exner originally included only scoring categories that had interscorer reliabilities of .85 or higher. Some controversy has resulted concerning these values in that other researchers have reported greater variability. Parker (1983) analyzed 39 papers using 530 different statistical procedures published in the *Journal of Personality Assessment* between 1971 and 1980. He concluded that, overall, the Rorschach can be expected to have reliabilities in the low to middle .80s. However, only two of his studies used the Comprehensive System. Acklin, McDowell, Verschell, and Chan (2000) found that nearly half of the categories for the Comprehensive System showed excellent reliabilities (>.81) with substantial reliability (.61 to .80) for a third of the categories. They concluded that a majority of the categories had excellent interscorer reliability, but a subset of about a quarter of the variables demonstrated less than adequate (<.61) reliability. The problem with the Acklin et al. data, however, was that the sample sizes were small, with the result that greater variability would be expected. In the most ambitious, rigorous, and large-scale study to date, G. Meyer et al. (2002) used eight different data sets and employed several different strategies to determine the reliability of the categories for the Comprehensive System. They concluded that it had overall excellent interscorer reliabilities with median correlations ranging from .82 to .97, depending on the data set used. Exner (2003) has recently reported new interscorer reliabilities with agreement ranging from a high of 99% for Texture and Vista responses to a low of 88% for passive movement. These correlations support Exner's claims that, if scorers are appropriately trained, the system has excellent interscorer reliabilities.

Test-retest reliabilities for the Comprehensive System have been somewhat variable. Retesting of 41 variables over a one-year interval for a nonpatient group produced reliabilities ranging between .26 and .92 (see Exner, 2003 in Table 11.3). Four of the correlations were above .90, 25 were between .81 and .89, and 10 were below .75. Exner has clarified that the 10 variables below .75 would all be expected to have had relatively low reliabilities because they related to changeable state (rather than trait) characteristics of the person. He also pointed out that the most important elements in interpretation are the ratios and percentages, all of which were among the higher reliabilities. Retesting for the same group over a three-year interval produced a similar but slightly lower pattern of reliability. In contrast, another group of nonpatient adults, retested over a much shorter (three-week) interval, had somewhat higher overall reliabilities than for either the one-year or three-year retestings (Exner, 1986). A more extensive summary of test-retest reliability by G. Meyer and Archer (2001) found that the mean reliability was .66 (range from .46 to .84, $Mdn = .69$). This is similar to the .66 to .82 mean reliabilities summarized by Viglione and Hilsenroth (2001). One issue, however, is that the Rorschach has approximately 125 variables, and some of these do not

have known test-retest reliability on them. The number of these untested reliabilities varies across researchers, with Wood and Lilienfeld (1999) stating that 85 variables have missing reliabilities and Viglione and Hilsenroth stating that only 12 variables have unknown test-retest reliabilities.

Long-term retesting for children has not come close to the same degree of stability as for adults (Exner, 2003; Exner & Weiner, 1995). Exner (1986) clarifies that this low stability for test results is to be expected, given that children undergo considerable developmental changes. However, short-term retesting over seven-day (for 8-year-olds) and three-week (for 9-year-olds) intervals did indicate acceptable levels of stability (Exner, 2003). Only 2 of 25 variables were below .70, with at least 7 above .90, and the remainder from .70 to .90. As with adults, the ratios and percentages demonstrated relatively high stabilities. Although acceptable short-term stability for young children's Rorschach variables was demonstrated, long-term stability was not found to occur until children reached the age of 14 years or older (Exner, 2003; Exner, Thomas, & Mason, 1985).

The primary focus of early validity studies was to discriminate empirically among different populations. This was based on past observations of a particular group's responses to the Rorschach, the development of norms based on these responses, and comparisons of an individual's Rorschach responses with these norms. For example, a person with schizophrenia might have a relatively high number of poor-quality responses, or a depressed person might have very few human movement responses. In addition to these empirical discriminations, efforts have been made to develop a conceptual basis for specific responses or response patterns (Weiner, 1994). Thus, it has been conceptualized that people with schizophrenia have poor-quality responses because they do not perceive the world the way most people do; their perceptions are distorted and inaccurate, and their reality-testing is poor. A further approach, which was not extensively developed in the Comprehensive System (nor by Rorschach himself), was the validation of the latent meaning of symbolical content.

These very general approaches have given rise to a surprisingly large number of specific scorings and interpretations, all of which have had various degrees of support. Many of the early validity studies are difficult to evaluate because of the varying scoring systems and poor methodologies. In addition, most early studies depended on inadequate norms (especially for studies conducted on children, adolescents, and persons over 70). Test results might also have been significantly influenced by situational and interpersonal variables, such as seating, instructions, rapport, gender, and personality of the examiner (see review by Masling, 1992). It should then come as no surprise that, for every study supporting an interpretive hypothesis, there would often be another refuting the same hypothesis.

Establishing the validity of the Rorschach as a whole has been further complicated by the many scoring categories and quantitative formulas, each of which has varying levels of validity. Some interpretations have greater validity than others even in a specific category. For example, the number of human movement responses (M) has been used as an index of both creativity and fantasy. A review of the research by Exner (1993) indicates that *M* relates fairly clearly to fantasy in that it has been correlated with daydreaming, sleep/dream deprivation, dream recall, and total time spent dreaming, whereas associations between *M* and creativity have been weaker and more controversial. Validity might

also depend on the context and population for which the test is used. For example, a depression index (DEPI) based on seven Rorschach combinations of scores has been found to provide moderate to low associations with the presence of depression among adults (Jorgenson, Anderson, & Dam, 2000; G. Meyer, 2000). However, among adolescent populations, the depression index was not successful in distinguishing depressives from schizophrenics (Archer & Gordon, 1988; Archer & Krishnamurthy, 1997a; Ball, Archer, Gordon, & French, 1991). In contrast to DEPI, an index designed to detect thought disorders (Schizophrenia Index or SCZI) has been somewhat more successful (Jorgenson et al., 2000; G. Meyer, 2000). Additional validity data on specific scoring categories and formulas are included in the Interpretation section of this chapter. These data should be carefully read to more fully understand Rorschach validity.

Probably the best way to provide a global index of validity is to combine the results from a large number of studies. Early meta-analyses indicated that validity ranged from .40 to .50 (L. Atkinson, 1986; L. Atkinson, Quarington, Alp, & Cyr, 1986; Parker, 1983; Parker, Hanson, & Hunsley, 1988; Weiner, 1996). However, these results have been challenged by Garb, Florio, and Grove (1998; Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Hunsley & Bailey, 2001) who reanalyzed the data from Parker et al. and concluded that the overall validity coefficients for the Rorschach were only .29 (in contrast to the significantly higher validity of .48 for the MMPI). This produced lively debates in the literature regarding the most appropriate methods of analysis. The majority of recent meta-analyses have continued to support the validity of the Rorschach (R. F. Bornstein, 1999; G. Meyer & Archer, 2001; G. Meyer & Handler, 1997). However, interactions with type of scoring system, experience of the scorer, and type of population used were likely to have complicated the picture.

One of the main efforts on establishing Rorschach validity has been directed toward determining its ability to discriminate among different types of populations. The success of these differentiations has been somewhat equivocal (see Vincent & Harman, 1991; Wood et al., 2000). For example, Wood et al. have indicated that, with the exception of a few disorders (schizophrenia, borderline personality, bipolar disorder), the Rorschach has not been very effective at assisting with making formal psychiatric diagnosis. A defense is that, in contrast to structured interviews or tests such as the MCMI-III, the Rorschach was not designed to accomplish this goal. This means that validity is conditional in that it is optimally valid for some purposes but not for others. Although the Rorschach is not the optimal instrument for most forms of diagnosis, it has been found to effectively predict variables such as outcome from psychotherapy (using the Prognostic Rating Scale; $r = .45$), detection of psychosis (using the Schizophrenia Index; $r = .44$), and dependent behavior (using the Oral Dependency Scale; $r = .37$; G. Meyer & Archer, 2001).

One major factor that may serve to lower Rorschach validity is the meaning associated with, and the effects of, response productivity. Various interpretations have been associated with extremes of productivity, with low productivity suggesting defensiveness, depression, and malingering, and extremely high productivity suggesting high achievement or an obsessive-compulsive personality. However, response productivity has also been found to be closely tied to age, intellectual level, verbal aptitude, and amount of education. Norms have been provided for different ages (Exner, 1993; Exner & Weiner, 1995), which can be helpful in correcting for the effects of age. However,

intellectual level, verbal aptitude, and amount of education can potentially confound the meanings associated with response productivity. A high number of responses does not necessarily represent traditional personality interpretations (obsessiveness, creativity, good impulse control), but might merely indicate a high level of verbal aptitude. Most early validity studies rarely considered these factors. More importantly, the number of responses not only affects interpretations related specifically to response productivity, but also productivity affects many other areas of interpretation. For example, a low number of responses is likely to increase the relative number of responses based on the whole inkblot (W). In contrast, a high number of responses would be likely to increase the relative number of small detail (Dd) responses. Because interpretations are frequently based on the relative proportions of different scoring categories (calculated in quantitative formulas), the overall number of responses is likely to influence and possibly compromise the validity of the formulas. However, Exner (1993) has found that lengthy records generally did not result in different interpretations when compared with records from the same persons with average numbers of responses. For practical reasons, he has recommended that the number of responses be limited if the person gives six or more responses to the first card or five or more responses to the second card (see Administration section). In contrast to lengthy protocols are ones with extremely low numbers of responses. Exner (2003) recommends that brief protocols (fewer than 14) be discarded and the test be readministered. This problem with the meaning of various numbers of responses largely led Holtzman to develop his alternate test (Holtzman Inkblot Test), in which subjects provide only one response for each inkblot in his series (Holtzman, 1988).

A further area of difficulty in establishing validity is that Exner cites extensive validity studies throughout his three volumes, but many of these studies were not done using his Comprehensive System. Comparability between the different studies and systems is frequently assumed or at least implied. However, these studies were frequently done at a time when norms were inadequate, interscorer reliability was questionable, and little concern was given to the possible confounding effects of age, intellectual level, education, and verbal aptitude. The development of the Comprehensive System itself was largely motivated by the deficiencies (and strengths) inherent in each of the earlier systems. More recently, there have been a greater proportion of studies that used the Comprehensive System, which has helped to reduce this problem. Eventually, these newer studies based on the Comprehensive System will help clarify Rorschach validity without the possible contaminating effects of previous work that used other systems.

The previous overview of Rorschach reliability and validity suggests a number of conclusions. Interscorer and test-retest reliabilities for the Comprehensive System have generally been supported although there are a number of variables with unknown test-retest reliability. The overall validity of the Rorschach has been found to be moderate to low (.30). What is less clear is which of the many variables are the most, versus the least, valid. For example, the Schizophrenia Index and Prognostic Rating Scale have been found to predict relevant external behaviors. In contrast, the Depression Index has clearly not been found to be effective with adolescents and children and is only minimally effective with adults. As a result, it should probably be dropped from the Comprehensive System. Future research will provide more clear conclusions related to the validity of the multitude of categories.

ASSETS AND LIMITATIONS

As mentioned previously, the Rorschach has been surrounded by controversy. Often, battle lines have been polarized into either "clinical loyalists" or "academic iconoclasts" (Parker, 1983). Despite thousands of research studies, these positions have changed only minimally over the past 60 years. L. Atkinson (1986) has suggested that the controversial status of the Rorschach may be largely the result of sociocultural factors rather than actual scientific evidence. It is hoped that the Comprehensive System, along with reviews such as G. Meyer and Archer (2001), will eventually represent a middle ground that will satisfy hard-nosed empiricists and address areas relevant to clinicians.

Part of the reason the Rorschach has continued to have such high popularity is the number of attractive features associated with it. Perhaps part of its allure is the mystery it frequently seems to invoke. How could something as seemingly simple as 10 inkblots reveal inner aspects of a person's personality? Metaphors such as "X rays of the mind" have certainly served to enhance its mystery and power. Often, a Rorschach protocol is perceived as something like a deep well, in that the skilled clinician can dip into it repeatedly, continually coming up with rich and valuable information. The practitioner is framed as a seer and an artist rather than a technician. Indeed, studies tend to support the belief that highly trained Rorschach experts can accurately describe a person's characteristics based on Rorschach responses. However, this accuracy has often been dependent more on intuition and clinical lore than on clearly validated interpretive rules.

One frequently noted asset is that the Rorschach is considered to be excellent at bypassing a person's conscious resistance; instead, it assesses a person's underlying, unconscious structure of personality. This asset might be particularly important if a person appears to have an adequate surface level of adjustment, yet the clinician suspects there may be some underlying pathology. In contrast, a structured test, such as the MMPI, may have difficulty assessing these more hidden levels of pathology. It is precisely the difficulty in organizing the ambiguous Rorschach stimuli that is likely to bring out these latent levels of pathology. There is some support for this view in that persons with borderline psychopathology have relatively normal performance on structured tests. In contrast, they tend to show clear indications of thought disorder on the far less structured Rorschach (Edell, 1987). Similarly, a relatively hidden trait such as alexithymia has been found in psychosomatic patients, based on their Rorschach responses (Acklin & Bernat, 1987; Keltikangas-Jarvinen, 1986). G. Frank (1990) reviewed the existing literature and found that the Rorschach was sensitive to underlying schizophrenic processes even before their clinical expression.

A related asset is the Rorschach's purported high resistance to faking. It is argued that, because the true meanings of the Rorschach responses are unknown, the subject cannot easily invent faked responses. Some proponents have even stated that it is virtually impossible to fake a Rorschach. Like many other statements about the Rorschach, this has become quite controversial. Exner (1993, 2003) has presented material, from a theoretical and empirical perspective, suggesting that persons developing a Rorschach response go through a series of six stages, one of which is censorship. Subjects seem to come up with far more responses than they present to the examiner, and they select the ones they feel are most appropriate to reveal. Subjects who feel emotionally close to the

examiner tend to provide more responses and conceal less (Leura & Exner, 1978). This raises the possibility that they might also have enough control over their responses to effectively fake a protocol. Thus, responses might depend to a certain extent on social desirability, perceptual accuracy, the context of the assessment, and personal needs. Despite the possibility of censorship, which might potentially lead to undetected faking, Exner and Wylie (1975) have reported that only 1 student in 12 could simulate a schizophrenic profile, even though the students were familiar with protocols from actual schizophrenics. Specifically, malingerers were likely to have longer free associations (presumably because they were censoring and elaborating on their responses), relatively accurate perceptions, and highly dramatic and idiosyncratic responses (i.e., "That's too awful to look at"). Similarly, Frueh and Kinder (1994) found that persons who were malingering with posttraumatic stress disorder provided responses that were overly dramatic, relatively unrestrained, and indicative of an exaggerated sense of impaired reality testing. In contrast, Albert, Fox, and Kahn (1980) found that Rorschach experts did poorly when requested to blindly classify protocols from normals who were requested to fake paranoid schizophrenia, normals taking a standard administration, and diagnosed paranoid schizophrenics. Computer analyses of the same protocols were likewise unsuccessful in effectively detecting faking (M. W. Kahn, Fox, & Rhode, 1988). Although this clearly challenges the unfakability of the Rorschach, the Albert et al. and Kahn et al. studies did not simulate the manner in which the Rorschach is likely to be used in clinical practice. Typically, practitioners have knowledge regarding the history of the person, context of the assessment, and behavioral observations, all of which potentially sensitize them to the possibility that a protocol might be faked. Consistent with this was the Frueh and Kinder study, which found that relevant behavioral observations were at least as important in detecting malingering as the actual scored protocols.

One clear asset of the Rorschach is its ease of administration. The cards can be easily handled, and the total administration time (including inquiry) is typically 50 minutes (Ball, Archer, & Imhof, 1994). In contrast to the relative ease of administration, scoring and interpretation are often quite complicated and time-consuming. Clinicians report that scoring usually takes 45 minutes and interpretation requires 50 minutes more (Ball et al., 1994). This means that, collectively, the entire procedure takes nearly 2.5 hours. However, computer-assisted scoring and interpretation would be expected to reduce significantly the time for both scoring and interpretation.

Alongside the advantages associated with the Rorschach, it has a number of limitations. Although both reliability and validity have generally reached adequate levels, validity is often quite variable across different scoring categories and formulas. Typically, multiple scores and formulas are derived from the Rorschach responses, some of which have relatively good validity and some of which are moderate, controversial, or even nonexistent. It is usually difficult for the average user to appreciate and take into account the disparate levels of validity when actually making his or her interpretations.

Because the Rorschach is one of the most complex psychological tests in current use, error can potentially be introduced from many different directions, including censorship by the subject, scoring errors (particularly for infrequently used scorings), poor handling of the subtleties of interpretation, incorrect incorporation of the implications of age or education, or possible examiner bias (illusory correlation, primacy effects, and so on). One temptation is to reduce the complexity of the data by using a

single-sign approach rather than viewing each sign in the context of the overall configuration. Rorschach "elevations" are often subject to a number of possible interpretive hypotheses, so a single-sign approach is particularly open to error. Thus, interpretations must be continually checked and rechecked against the overall Rorschach configuration, additional test data, and the patient's history.

The complexity of the Rorschach also requires that potential users undergo extensive training. Each new scoring category and index that is introduced may add to this problem. In the past, graduate schools would sometimes provide a full-semester course on the Rorschach. Some authors, feeling that this is insufficient, have stated that the optimum amount of time is two full-semester courses devoted exclusively to the Rorschach (Hilsenroth & Handler, 1995), a curriculum that is currently difficult for many programs to justify. First, other tests often have superior psychometric properties. Second, the past 25 years have brought a significant increase in the roles and skills required of graduate students, including skills in the area of assessment (neuropsychology, behavioral assessment) as well as in other areas of clinical practice (family therapy, rehabilitation, new modes of intervention, treatment of chronic pain, and so on). Despite these increased requirements, the many training programs continue to expect trainees to have or to develop skills in the Rorschach (Watkins et al., 1995).

A further difficulty associated with the Rorschach is the previous lack of a single, standardized administration and scoring system. This is particularly important because numerous studies have clearly indicated that slight alterations in wording, rapport, and encouragement can significantly alter the numbers and types of responses. The numerous differences in administration and scoring will, it is hoped, be seen in the future as a historical aberration and will be corrected by wide acceptance of the Comprehensive System. However, belief that a single unified system will be accepted should be tempered by recent reviews indicating that between 25% and 33% of Rorschach courses do not use the Comprehensive System (Hilsenroth & Handler, 1995).

The Rorschach has often been considered to have limited use with children, particularly children under the age of 14 years (Klein, 1986). Reliabilities have been found to be adequate for short-term assessments but clearly inadequate over a long-term basis. Thus, for purposes such as child custody decisions, where longer term predictions are required, the Rorschach would be quite limited. Any use of the Rorschach for children should make clear that descriptions are only for the short term.

A final consideration, which has implications for both research and practice, is that the large number of variables is likely to produce spurious random significance. Wechsler subtest interpretation has attempted to correct for this possibility by carefully calculating the significance of subtest differences, including correction factors for the number and reliabilities of variables considered (see Chapter 5). In contrast, it is difficult to know when the numerous variables considered in the Rorschach might indicate "significance" simply because of random fluctuations of scores (i.e., a .05 level of significance would mean that "significance" would happen by chance in 1 of 20 variables considered). Rorschach interpreters must, therefore, take extra caution with their interpretations.

In summary, the Rorschach is difficult to evaluate because of its complexity, its frequent controversy, and considerable variability related to the validity of its variables. The voluminous research associated with the Rorschach is often both an asset and a

limitation. Sorting through the maze of sometimes contradictory findings is often difficult. Directing this wealth of research toward a clear understanding of the interpretive meanings associated with certain patterns of scores is especially difficult. The specific assets of the Rorschach are potential wealth of information, simplicity of handling, ability to bypass conscious resistance, and possible resistance to faking. Significant weaknesses are moderate and sometimes quite variable reliabilities and validities, time required for scoring and interpretation, limited use with children, extensive time required for training, and possible introduction of error, especially spurious random significance as a result of the large number of areas considered.

APPROACHING CONSENSUS ON RORSCHACH VALIDITY

Given the previous review and controversy over the Rorschach, an important priority is to develop a consensus on those categories that are best validated compared to those that are questionable (see, for example, how this was done with intelligence in Neisser et al., 1996, "Intelligence: Knowns and Unknowns"). This would help to refine interpretive systems as well as provide clear guidance for practitioners. One difficulty in this process is the volumes of research literature that would need to be evaluated and integrated. Another difficulty is the extent to which opposing sides are polarized. The outcome would be a listing of all categories that are clearly supported, an intermediate grouping on which there is either disagreement or equivocal conclusions, and a final listing of all categories that clearly have insufficient support and should not be used.

It is hoped that the following summary is the beginning of such a process. This is necessarily a very small portion of what needs to be done because there is so much to cover. There are also a very large number of categories that are not covered. Even the staunchest Rorschach critics indicate that at least some of the Rorschach categories are well supported. Wood and Lilienfeld (1999) conclude that schizophrenics, persons with schizotypal personality disorder, and bipolar patients provide more poor form responses and Deviant Verbalizations (this is also reflected in indexes that use these categories). Borderlines also provide more Deviant Verbalizations. Garb (personal communication) also agrees with these findings on Deviant Verbalizations and endorses the Rorschach Prognostic Rating Scale and the Rorschach Oral Dependency Scale. Rorschach proponents would obviously agree with this as well as include a large number of additional categories (many of which are detailed with relevant research later in the interpretation of this chapter).

There are also a number of Rorschach variables that have questionable reliability. Although most scoring categories have good to excellent interrater agreement, those under .75 include Texture Form (.54), Color' Form (.62), Form Quality+ (.66), Color' (.71), Fictional Animal detail (.72), and Incongruous Combination (.74; G. Meyer et al., 2002). Variables with low (under .70) test-retest reliability (one-year duration) include Inanimate Movement (.26), Diffuse Shading (.31), Pure Color (.56), Color Form (.58), Blends (.62), and Experienced Stimulation (.64; Exner, 2003). It is, however, arguable that some of these (i.e., inanimate movement, experienced stimulation) would, as Exner has pointed out, be expected to have low test-retest reliability because they are changeable characteristics of the person.

A number of variables with questionable validity have also been identified. For example, Archer and Krishnamurthy (2003) concluded that weak variables include the Depression Index (DEPI), Schizophrenia Index (SCZI), and Suicide Constellation (S-CON). In addition, relatively new variables that should be researched further include Perceptual Thought Index (PTI; this has replaced the Schizophrenia Index or SCZI), Coping Deficit Index (CDI), Hypervigilance Index (HVI), Obsessive Style Index (OBS), Good Human Representation (GSR), Poor Human Representation (PHR), Form Appropriate-Extended (XA%), and Form Appropriate-Common Areas (WDA%). As a result, these should be interpreted with considerable caution. The Rorschach also should not be used for the assessment of brain damage or identification of sexual abuse; and there are certainly better instruments available to develop formal *DSM-IV* (1994) diagnoses. No doubt additional variables will be identified and, if indeed found to be weak, will be excluded. The result would be a shorter but also well-validated instrument that would be accepted even by its critics. It is hoped that this process continues such that the polarity in the Rorschach debates will be reduced and clinicians can improve the accuracy of their interpretations, thereby improving client care.

ADMINISTRATION

Examiners should standardize their administration procedures as much as possible. This is particularly important because research has consistently indicated that it is relatively easy to influence a subject's responses. For example, saying the word *good* after each response can increase the overall number of responses on the Rorschach by as much as 50% (Hersen & Greaves, 1971). Similarly, examiners who were told that more experienced examiners elicited a greater proportion of human than animal responses actually produced this pattern from examinees, even though the examiners believed they were providing a standard administration (Exner, Leura, & George, 1976). These findings are consistent with the view that subjects are particularly responsive to subtle influences when attempting to create clarity in an ambiguous situation such as projective testing. However, if the fluctuations in administration style are minor, they are unlikely to significantly influence a subject's responses. In general, examiners should minimize the variations in their administration procedures as much as possible. The following sequence of steps is recommended.

Step 1: Introducing the Respondent to the Technique

One of the most important goals an examiner must initially achieve is to allow the examinee to feel relatively comfortable with the testing procedure. Achieving this goal is complicated by the fact that tests in most cultures are associated with anxiety. Although, in some cases, an increase in anxiety may provide some information that cannot be obtained when the subject is relaxed, anxiety is usually regarded as a hindrance. Typically, anxiety interferes with a person's perceptions and with the free flow of fantasy, both of which are essential for adequate Rorschach responses. Thus, subjects should be as relaxed as possible. Their relaxation can be enhanced by giving a clear introduction to the testing procedure, obtaining personal history, answering questions,

and generally avoiding any behavior that might increase the subjects' anxiety. In describing the test, examiners should emphasize relatively neutral words such as *inkblot, interests,* or *imagination,* rather than potentially anxiety-provoking words such as *intelligence* or *ambiguous.*

For the most part, any specific information regarding what subjects should do or say is to be avoided. The test situation is designed to be ambiguous, and examiners should avoid any statements that might influence the responses. If subjects push for more detailed information about what they should do or what their responses may mean, they should be told that additional questions can be answered after the test is completed.

Step 2: Giving the Testing Instructions

Although some Rorschach systematizers recommend that the subject tell the examiner "everything you see" (S. J. Beck, 1961), the Comprehensive System attempts to keep the task as ambiguous as possible. Thus, Exner (1993) recommends that the examiner hand the subject the first card and ask, "What might this be?"

Commentary on, or discussion of the cards by the examiner, should be avoided as much as possible. At times, it might be acceptable to briefly describe how the designs were made, or, if questioned regarding what one is supposed to see, the examiner might state, "People see all sorts of things in the blots." Comments from the examiner that indicate the quantity or type of response, or whether the subject can turn the cards, should be strictly avoided. If the subject asks specific questions, such as the type of responses he or she is supposed to give or whether he or she can turn the cards, the examiner might reply that it is up to him or her to decide.

The main objective is to give the subject maximum freedom to respond to the stimuli in his or her own manner. To enhance this, Exner (2003) strongly recommends that the subject and the examiner not be seated face-to-face, but rather side-by-side, to decrease the possible influence of the examiner's nonverbal behavior. The overall instructions and testing situation should be designed both to keep the task as ambiguous as possible and to keep examiner influence to a minimum.

Step 3: The Response (Association) Phase

Throughout the testing procedure, the basic conditions of step 2 should be adhered to as closely as possible. However, specific situations often arise as subjects are free-associating to the Rorschach designs. If a subject requests specifics on how to respond or asks the examiner for encouragement or approval, examiners should consistently reply that he or she can respond however he or she likes. The idea that there are no right or wrong answers might sometimes be mentioned.

The examiner should time the interval that begins when subjects first see the card and ends when they make their initial response, as well as the total time they spend with each card. These measurements can be helpful in revealing the general approach to the card and the possible difficulties in coming up with responses. Cards II, III, and V are generally considered relatively easy to respond to, and usually have shorter reaction times than Cards VI, IX, and X, which typically produce the longest reaction times. Because overt timing of subjects' responses is likely to produce anxiety, any recording

should be done as inconspicuously as possible. It is recommended that, rather than using a stopwatch, the examiner glance at a watch or clock and record the minute and second positions for the initial presentation, the first response, and the point at which the subject hands the card back to the examiner.

The average number of responses is 22.32 (average range = 17 to 27). Validity can be compromised with a low number of responses (under 14) and may be questionable with a high number of responses (more than 42). Exner (2003) has built in some safeguards to protect against unusually short or extremely long protocols. A client who produces an extremely brief protocol (fewer than fourteen responses) should be immediately retested and provided with a clearer request to provide more responses (Exner, 2003). If a client provides six or more responses to the first inkblot, the examiner should remove the inkblot. On all subsequent inkblots, the same procedure should be used whenever the client provides five or more responses. However, if fewer than six responses to the first inkblot are given, no other limits on either the first inkblot or any later inkblots should be provided.

Exner (2003) stresses that all responses must be recorded verbatim. To simplify this process, most clinicians develop a series of abbreviations. A set of abbreviations used throughout all the Rorschach systems consists of the symbols (V, > , Λ, <) in which the peak indicates the angle of the card. It is also important to note any odd or unusual responses to the cards, such as an apparent increase in anxiety, wandering of attention, or acting-out on any of the percepts.

Step 4: Inquiry

The inquiry should begin after all 10 cards have been administered. Its purpose is to collect the additional information required for an accurate scoring of the responses. It is intended to clarify the responses that have already been given, not to obtain new responses. The inquiry should not end until this goal has been accomplished. Exner (2003) recommends that the instructions for the inquiry closely approximate the following:

> Now we are going to go back through the cards again. It won't take very long. I want to see the thing that you said you saw and make sure that I see them like you do. We'll do them one at a time. I'll read what you said and then I want you to show me where it is in the blot and then tell me what there is there that makes it look like that to you, so that I can see it too, just like you did. Is that clear?

Following closely the general theme of the overall administration, the inquiry should not influence the subject's responses. Thus, any questions should be as nondirective as possible. The examiner should begin by merely repeating what the subject has said, and then waiting. Usually, the subject begins to clarify his or her response. If this information is insufficient to clarify how to score the response, the examiner might become slightly more directive by asking, "What about it made it look like a [percept]?" The examiner should not ask, "Is it mainly the shape?" or "How important was the color?" These questions are far too directive and are worded in a way that can exert influence on the subject's descriptions of his or her responses. The examiner should consistently avoid leading the subject or indicating how he or she should respond. Particular

skill is required when clarifying a determinant that has been unclearly articulated but merely implied.

The outcome of a well-conducted inquiry is the collection of information sufficient to decide on scoring for location and determinants. If, on the location, information based on the subject's verbal response is insufficient, the examiner should have the subject point to the percept. An additional feature of the inquiry is to test the subject's awareness of his or her responses. For example, does a strange percept represent coherent creativity, or does it reflect a lack of contact with the environment, with the subject perhaps having no awareness of the strangeness of his or her responses? The overall approach of the inquiry is to word questions in such a way as to be flexible without being too directive.

SCORING

The next step following administration is to code the different categories and calculate the different quantitative formulas in the structural summary. There is general agreement throughout the different Rorschach systems that these categories include location, determinants, content, and popularity. The Comprehensive System also includes 15 special scores for responses such as unusual verbalizations and aggressive movement. After these have been coded and tallied, a series of quantitative summaries, including six Special Indices, are created based on reorganizations of, and comparisons among, the scores on the different categories.

The following subsections merely list, outline, and define the scoring categories and quantitative summaries. To achieve accurate scoring, it would be necessary to consult Exner's scoring criteria (2003) or to implement his workbook (Exner, 2001, *A Rorschach Workbook for the Comprehensive System,* 5th ed.), which includes specific scoring criteria, tables, charts, and diagrams. The inclusion of specific scoring criteria is beyond the scope of this chapter. The focus here is on providing a key to interpretation that is concise, accountable, and clearly organized. The following definitions and the accompanying tables serve to outline and briefly define the primary Rorschach factors.

Table 10.1 Symbols used for coding the location of Rorschach responses

Symbol	Definition	Criterion
W	Whole response	Where the entire blot is used in the response. All portions must be used.
D	Common detail response	A frequently identified area of the blot.
Dd	Unusual detail response	An infrequently identified area of the blot.
S	Space response	A white-space area is used in the response (scored only with another location symbol, as in WS, DS, DdS).

Source: From *The Rorschach: A Comprehensive System. Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, New York: John Wiley & Sons. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Location

The *location* of the responses refers to the area of the inkblot that is used (Table 10.1). This can vary from the use of the entire blot (whole response) to the use of an unusual detail (Dd). Unusual details are defined as location responses made by less than 5% of subjects. Exner also specifies coding for Developmental Quality, which is determined by evaluating each location score in relation to its degree of integration. Table 10.2 presents the criteria used for scoring the respective Developmental Quality codes. Thus, each location response is given both a designation for the specific area of the blot and a symbol to indicate the quality of that response.

Determinants

Determinants refer to the style or characteristic of the blot to which the examinee responds, such as its shape, color, or texture (Table 10.3). The determinants also receive a scoring for their level of form quality (Table 10.4). The form quality scoring refers to how accurately the percept relates to the form of the inkblot. For example, an angel on Card I is considered to be an "ordinary" form quality response, which is empirically

Table 10.2 Symbols and criteria used for developmental quality

Symbol	Definition	Criterion
+	Synthesized response	Two or more objects are described as separate but related. *At least one* of the objects involved must have a specific form demand, or be described in a manner that creates a specific form demand (e.g., a dog walking among some bushes, a man with a funny hat on, an airplane flying through some clouds, the head of a little girl, she has a hair ribbon).
o	Ordinary response	An area of the blot is identified as a single object that has features that create a natural form demand or the *description of the object is such as to create* a specific form demand (e.g., a fir tree, a cat, a totem pole, a maple leaf, a bat, a flag, a man's head).
v/+	Synthesized response	Two or more objects are described as separate but related. *None of the objects* involved have a specific form demand and the articulation does not introduce a form demand for any of the objects (e.g., clouds coming together, some sort of bay with the vegetation around the shore, a rock and some dirt around it).
v	Vague response	An object is reported that has no specific form demand, *and the articulation does to introduce* a specific form demand for the object (e.g., a cloud, the sky, the colors of sunset, some ice).

Source: From *The Rorschach: A Comprehensive System. Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, New York: John Wiley & Sons. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Table 10.3 Symbols and criteria for determinant coding

Category	Symbol	Criterion
Form	*F*	*Form answers.* Used for responses based exclusively on the form features of the blot.
Movement	*M*	*Human movement response.* Used for responses involving the kinesthetic activity of a human or an animal or fictional character in human-like activity.
	FM	*Animal movement response.* Used for responses involving a kinesthetic activity of an animal. The movement perceived must be congruent to the species identified in the content. Animals reported in movement not common to their species should be coded as *M*.
	m	*Inanimate movement response.* Used for responses involving the movement of inanimate, inorganic, or insensate objects.
Chromatic color	*C*	*Pure color response.* Used for answers based exclusively on the chromatic color features of the blot. *No* form is involved.
	CF	*Color-form response.* Used for answers that are formulated *primarily* because of the chromatic color features of the blot. Form features are used but are of secondary importance.
	FC	*Form-color response.* Used for answers that are created mainly because of form features. Chromatic color is used but is of secondary importance.
	Cn	*Color naming response.* Used when the colors of the blot are identified *by name* and with the intention of giving a response.
Achromatic color	*C′*	*Pure achromatic color response.* Used when the response is based exclusively on the grey, black, or white features of the blot, when they are clearly used as color. *No* form is involved.
	C′F	*Achromatic color-form response.* Used for responses that are created *mainly* because of the black, white, or grey features, clearly used as color. Form features are used but are of secondary importance.
	FC′	*Form-achromatic color response.* Used for answers that are based *mainly* on the form features. The achromatic features, clearly used as color, are also included but are of secondary importance.
Shading-texture	*T*	*Pure texture response.* Used for answers in which the shading components of the blot are translated to represent a tactual phenomenon, with no consideration to the form features.
	TF	*Texture-form response.* Used for responses in which the shading features of the blot are interpreted as tactual, and form is used secondarily, for purposes of elaboration and/or clarification.
	FT	*Form-texture response.* Used for responses that are based *mainly* on the form features. Shading features of the blot are translated as tactual but are of secondary importance.

(continued)

Table 10.3 *(Continued)*

Category	Symbol	Criterion
Shading-dimension	*V*	*Pure vista response.* Used for answers in which the shading features are interpreted as depth or dimensionality. *No* form is involved.
	VF	*Vista-form response.* Used for responses in which the shading features are interpreted as depth or dimensionality. Form features are included but are of secondary importance.
	FV	*Form-vista response.* Used for answers that are based mainly on the form features of the blot. Shading features are also interpreted to note depth and/or dimensionality but are of secondary importance to the formulation of the answer.
Shading-diffuse	*Y*	*Pure shading response.* Used for responses that are based exclusively on the light-dark features of the blot that are completely formless and do not involve reference to either texture or dimension.
	YF	*Shading form response.* Used for responses based primarily on the light-dark features of the blot, not involving texture or dimension. Form features are included but are of secondary importance.
	FY	*Form-shading response.* Used for responses that are based mainly on the form features of the blot. The light-dark features of the figure, not used to articulate texture or dimension, are included as elaboration and/or clarification and are secondary to the use of form.
Form dimension	*FD*	*Form-based dimensional response.* Used for answers in which the impression of depth, distance, or dimensionality is created by using the elements of size and/or shape of contours. *No* use of shading is involved in creating this impression.
Pairs and reflections	*(2)*	*The pair response.* Used for answers in which two identical objects are reported, based on the symmetry of the blot. The objects must be equivalent in all respects, but must *not* be identified as being reflected or a as mirror images.
	rF	*Reflection-form response.* Used for answers in which the blot or blot area is reported as a reflection or mirror image because of the symmetry of the blot. The object or content reported has no specific form requirement, as in clouds, landscape, shadows, and so on.
	Fr	*Form-reflection response.* Used for answers in which the blot or blot area is identified as reflected or a mirror image, based on the symmetry of the blot. The substance of the response is based on form features, and the object reported as a specific form demand.

Source: From *The Rorschach: A Comprehensive System. Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, New York: John Wiley & Sons. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Table 10.4 Symbols and criteria for coding form quality

Symbol	Definition	Criterion
+	Ordinary-elaborated	The unusually detailed articulation of *form* in responses that otherwise would be scored ordinary. It is done in a manner that tends to enrich the quality of the response without sacrificing the appropriateness of the form use. The + answer is not necessarily original or creative; but, rather, it stands out by the manner in which form details are used and specified.
o	Ordinary	The common response in which general form features are easily articulated to identify an object. These are easy-to-see answers that have been reported by at least 2% of persons in the Form Quality data pool for *W and D* areas, or a by at least 50 persons in the pool who responded to *Dd* areas. There is no unusual enrichment of the answer by elaboration of the form features.
u	Unusual	A low-frequency response in which the basic contours involved are appropriate for the response. These are uncommon answers that are seen quickly and easily by the observer.
–	Minus	The distorted, arbitrary, unrealistic use of form in creating a response. The answer is imposed on the blot structure with total, or near total disregard for the contours of the area used. Often, substantial arbitrary lines or contours will be created where none exist.

Source: From *The Rorschach: A Comprehensive System. Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, New York: John Wiley & Sons. Copyright © 2003 by John Exner Jr. Reprinted with permission.

reflected in the fact that nonpsychiatric populations perceive it far more frequently than psychiatric patients. Initially, examiners should give a percept its appropriate classification regarding its determinants. This should then be followed by scoring the determinant for its relative form quality. Descriptions of the different form qualities are included in Table 10.4; however, for specific empirically derived form quality codings, examiners need to consult Exner's (2003) tables.

One relevant coding that should be added to all movement responses is the extent to which the movement is active versus passive. Active movement would include movements such as "fleeing" or "lifting," whereas more passive movements might include "meditating" or "anchored." Whether a movement is active or passive is designated with either an *a* (for active) or a *p* (for passive) superscript. These are later scored and used for interpretation in the quantitative summaries (see the Ideation Section topic in the Structural Summary section).

In approximately 20% of all responses, more than one determinant is used to make a single response. These are referred to as *blends* and are designated by indicating the two (or more) determinants and placing a full stop (.) between them. The most important determinant is placed in front of the other determinant(s) and is considered the primary determinant. Less important determinants are placed after the primary one and are referred to as secondary or tertiary (if a third one is present).

A further score related exclusively to form determinants is the degree of Organizational Activity (Z) involved in creating the response. However, Organizational Activity is given only if at least one of the following four criteria is present:

1. A W response with DQ codings of +, o, v/+ (Wv responses are not scored for Organizational Activity).
2. The response gives some sort of meaningful integration to two or more areas (either adjacent or nonadjacent).
3. Two or more separate objects are identified in distant (nonadjacent) detail areas and these objects are described in some meaningful relationship.
4. The white space is given some sort of meaningful integration with other areas of the blot.

Specific converted weightings (ranging between 1 and 6) are given to integrative efforts for different types of responses and are provided in Exner (2003; see Table 8.4). For example, the degree of organization required to integrate a whole response to Card I is considered to be much less (Z would equal only 1.0) than that required to integrate the much more fragmented details of Card X (Z would equal a much greater 6.5).

Content

The scoring of content is based on the type and quantity of specific subjects that examinees perceive in their responses. Each Rorschach system uses different lists of content categories, although they all agree on basic contents such as human, human detail, and animal. Table 10.5 provides a listing of Exner's content categories, with the symbol and criterion for each category.

When two or more content categories occur in the same response, they should both be coded and a comma should be placed between the two (or more) codings. If contents occur that are not on the list, they should be designated as idiographic (Id) and the unique name of the content should be written out.

Popular Responses

Rorschach Popular (P) scoring refers to the presence of frequently perceived responses. Although different systems have somewhat varying lists of Populars, Exner (2003) has used, as the cutoff for inclusion as a Popular, an occurrence of at least once in every three protocols from nonpsychiatric populations. Exner's list of Popular responses is detailed in Table 10.6.

Special Scores

The Comprehensive System also includes 15 Special Scoring categories that were developed to take into account unusual characteristics of the response such as unusual verbalizations or inappropriate logic. These, along with their definitions, are listed in Table 10.7. A weighted sum of the first six categories (WSum6) is also required. The

Table 10.5 Symbols and criteria used for coding content

Category	Symbol	Criterion
Whole human	*H*	For responses involving a whole human form. If the response involves a *real* historical figure, such as Napoleon, Joan of Arc, and so on, the content code *AY* should be added as a secondary code.
Whole human, fictional or mythological	*(H)*	For responses involving a whole human form that is fictional or mythological, such as clowns, fairies, giants, witches, fairy-tale characters, angels, dwarfs, devils, ghosts, science-fiction creatures that are humanoid, human-like monsters, silhouettes of human figures.
Human detail	*Hd*	For responses involving an incomplete human form, such as an arm, head, leg, fingers, feet, the lower part of a person, a person without a head.
Human detail, fictional or mythological	*(Hd)*	For responses involving an incomplete human form that is fictional or mythological, such as the head of the devil, the arm of a witch, the eyes of an angel, parts of humanoid science-fiction creatures, jack-o-lantern, and masks *except* animal masks.
Human experience	*Hx*	Usually coded as a secondary content for answers that clearly involve the attribution of a human emotion or sensory experience to the object(s) in the response, such as *two people who are in love looking at each other, a cat that is very sad, people who are angry at each other, a woman smelling something nasty, a very happy person, a man who is very excited, a person in great pain.* The attribution of the motion or sensory experience must be clear and unequivocal. Answers such as *people at a party, an angry-looking face, a mean-looking person, two people who look tired* are not coded *Hx* as the attribution is equivocal. *Hx* is scored as a primary content for formless *M* responses that involve the emotion or sensory experience such as love, hate, depression, happiness, sound, smell, fear, and so on. These answers will also include the use of AB as a special score.
Whole animal	*A*	For responses involving a whole animal form.
Whole animal, fictional or mythological	*(A)*	For responses involving a whole animal that is fictional or mythological, such as a unicorn, dragon, magic frog, flying horse, Black Beauty, Jonathan Livingston Seagull.

(continued)

Table 10.5 *(Continued)*

Category	Symbol	Criterion
Animal detail	*Ad*	For responses involving an incomplete animal form, such as the hoof of a horse, claw of a lobster, head of a dog, animal skin.
Animal detail, fictional or mythological	*(Ad)*	For responses involving an incomplete animal form that is fictional or mythological, such as the wing of Pegasus, the head of Peter Rabbit, the legs of Pooh Bear, and all animal masks.
Anatomy	*An*	For responses in which the content is skeletal, muscular, or of internal anatomy such as bone structure, skull, rib cage, heart, lungs, stomach, liver, muscle fiber, vertebrae, brain. If the response involves a tissue slide, the content *Art* should be added as secondary.
Art	*Art*	For responses of paintings, drawings, or illustrations, either abstract or definite, art objects such as statues, jewelry, chandelier, candelabra, crests, badges, seals, and decorations. A feather seen worn as a decoration, often seen on Card VII, also should be coded as Art. In many responses coded for Art a second content will also be coded, such as a painting of two dogs would be *Art, A,* a sculpture of two witches would be *Art, (H),* a caricature of two people bending over would be *Art, H.*
Anthropology	*(Ay)*	For responses that have a specific cultural or historical connotation such as totem, roman helmet, Magna Carta, Santa Maria, Napoleon's hat, Cleopatra's crown, arrowhead, prehistoric axe, an Indian war bonnet.
Blood	*Bl*	For responses of blood, either human or animal.
Botany	*Bt*	For responses involving any plant life such as bushes, flowers, seaweed, trees or parts of plant life, such as leaves, petals, tree trunk, root, bird's nest.
Clothing	*Cg*	For responses involving any article of clothing such as hat, boots, belt, dress, necktie, jacket, trousers, scarf.
Clouds	*Cl*	For responses used specifically for the content cloud. Variations of this category, such as fog or mist are coded *Na.*
Explosion	*Ex*	For responses involving a blast or explosion, including fireworks.
Fire	*Fi*	For responses of fire or smoke.

(continued)

Table 10.5 *(Continued)*

Category	Symbol	Criterion
Food	*Fd*	For responses used for any edible common for humans, such as fried chicken, ice cream, fried shrimp, vegetables, cotton candy, chewing gum, steak, a filet of fish, or for animals eating a food that is natural for their species, such as a bird eating a worm or insect.
Geography	*Ge*	For responses used for the response of a map, specified or unspecified.
Household	*Hh*	For responses used for responses that include household items, such as bed, carving knife, chair, cooking utensils, cup, garden hose, glass, lamp, lawn chair, plate, rug *(animal skin rug should be coded Ad and Hh entered as a secondary content),* silverware. Some items coded *Hh* will also be coded as *Art,* such as candelabra, chandelier, or artistic pieces such as a centerpiece bowl.
Landscape	*Ls*	For responses that involve landscape, such as mountain, mountain range, hill, island, cave, rocks, desert, swamp, or seascapes, such as coral reef or underwater scene.
Nature	*Na*	For responses used for a broad variety of contents from the natural environment that are not coded as *Bt* or *Ls,* such as sun, moon, planet, sky, water, ocean, lake, river, ice, snow, rain, fog, mist, rainbow, storm, tornado, night, raindrop.
Science	*Sc*	For responses that are associated with, or are the direct or indirect products of science or science fiction, such as airplanes, buildings, bridges, cars, light bulb, microphone, motorcycles, motors, musical instrument, radar station, road, rocket ships, ships, space ships, trains, telescope, TV aerial, weapons, and so on.
Sex	*Sx*	For responses involving sex organs or activity of a sexual nature, such as penis, vagina, buttocks, breasts (except when used to identify the sex of a human figure), testes, menstruation, abortion, intercourse. *Sx* is usually scored as a secondary content. Primary contents are typically *H, Hd,* or *An.*
X-ray	*Xy*	For responses used specifically for the content of x-ray and may include either skeleton or organs. When *Xy* is coded, *An* is *not* included as a secondary code.

Source: From *The Rorschach: A Comprehensive System. Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, New York: John Wiley & Sons. Copyright © 2003 by John Exner Jr. Reprinted with permission.

Table 10.6 Popular responses used in the comprehensive system plus the proportions of each appearing in samples of nonpatient and patient protocols

Card	Location	Criterion	Nonpatient %	Patients %
I	*W*	Bat, with the true apex of the blot being identified as the top portion of the bat, and always involving the whole blot.	48	38
I	*W*	Butterfly, with the true apex of the blot being identified as the top portion of the butterfly, and always involving the whole blot.	40	36
II	*D1*	Animal, specifically identified as bear, dog, elephant, or lamb. The response is usually the head or upper body, but responses involving the whole animal are also coded *P*.	34	35
III	*D9*	Human figures or representations thereof such as dolls, caricatures, and so on. If *D1* is used as two human figures, *D7* or *Dd31* should not be reported as part of the human figure.	89	70
IV	*W* or *D7*	Human or human-like figure such as giant, monster, science-fiction creature, and so on. Animal figures are not coded as *P*.	53	41
V	*W*	Butterfly, with the true apex of the blot being identified as the top portion of the butterfly, and always involving the whole blot. The whole blot *must* be used.	46	43
V	W	Bat, with the true apex of the blot being identified as the top portion of the bat, and always involving the whole blot.	36	38
VI	*W* or *D1*	Animal skin, hide, rug, or pelt. Often, the skin, hide, or pelt will be included in the description of a whole animal, such as a cat or fox, in natural or unnatural form. The decision about whether to code *P* in these responses is based on whether the skin or hide is actually mentioned or clearly implied.	87	35
VII	*D1* or *D9*	Human head or face, specifically identified as female, child, Indian, or with gender not identified. This Popular is usually embedded in answers given to the larger areas, *D1, D2,* or *Dd23*. If *D1* is used, the upper segment (*D5*) is typically identified as hair or a feather. If the response includes the entire *D2* or *Dd23* areas, *P* is coded only if the head or face is restricted to the *D9* area.	59	47

Table 10.6 *(Continued)*

Card	Location	Criterion	Nonpatient %	Patients %
VIII	*D1*	Whole animal figure, usually of the canine, feline, or rodent varieties, with the head of the animal adjacent to the *D4* area.	94	91
IX	*D3*	Human or human-like figures such as witches, giants, science-fiction creatures, monsters, and so on.	54	24
X	*D1*	Spider with all appendages restricted to the *D1* area.	42	34
X	*D1*	Crab with all appendages restricted to the *D1* area. Other variations of multilegged animals are not *P*.	37	*38*

Source: From *The Rorschach: A Comprehensive System. Volume I: Basic Foundations* (4th ed.), by J. E. Exner Jr., 2003, New York: John Wiley & Sons. Copyright © 2003 by John Exner Jr. Reprinted with permission.

weightings are as follows: Deviant Verbalization (DV) = 1, Deviant Response (DR) = 3, Incongruous Combination (INCOM) = 2, Fabulized Combination (FABCOM) = 4, Inappropriate Logic (ALOG) = 5, and Contamination (CONTAM) = 7. A weighted score is given each time the scoring is given. For example, three occurrences of Deviant Response (DR) would equal a sum weighted score of 9.

In addition to the weighting, scoring according to levels are given for the severity of the first six Special Scores. A Level 1 score is considered to be fairly normal but a Level 2 score is considered to be a more pathological example of the response.

STRUCTURAL SUMMARY

After the examinee's responses have been coded according to locations, determinants, contents, Populars, and special scorings, they are listed and rearranged into frequency summaries and quantitative formulas. The quantitative formulas comprise various ratios, percentages, and derivations. These formulas reflect the proportions of, and comparisons among, various Rorschach factors. After the quantitative formulas have been calculated, they become the primary focus on which Rorschach interpretations are made. Exner (2003) has categorized the formulas into a Core section followed by sections for Ideation, Affect, Mediation, Processing, Interpersonal, Self-Perception, and Special Indices (Depression Index, Obsessive Style Index, and so on). These sections provide a convenient way to thematically organize the different interpretations. The descriptions and their sequence closely follow those outlined by Exner (2003). The various scorings, frequencies, and formulas can be conveniently summarized on a commercially available record form that includes a *Structural Summary Blank* as well as a *Constellation Worksheet* for calculating the Special Indices.

Table 10.7 Symbols and descriptions for special scores

Special Score (Symbol)	Description
Deviant Verbalization (DV)*	Verbalizations associated with a response, which are odd and suggest some form of cognitive slippage has occurred, such as through neologisms or redundancies (i.e., "pair of two").
Deviant Response (DR)*	Responses that involve a longer segment of the response than verbalizations, such as through inappropriate phrases or circumstantial responses that are long, rambling, and unrelated to the inkblot.
Incongruous Combination (INCOM)*	Images that have been inappropriately merged into a single object.
Fabulized Combination (FABCOM)*	Implausible relationships between two or more portions of the inkblot.
Contamination (CONTAM)*	Two or more impressions that have been inappropriately fused together.
Inappropriate Logic (ALOG)*	Spontaneously offered justification of the response using strained logic.
Perseveration (PSV)	Providing either an identical or almost identical response two or more times in a row, or seeing the same object repeatedly ("There's that man again").
Abstract Content (AB)	Symbolic representation is given to the content.
Aggressive Movement (AG)	Any movement response that is clearly aggressive.
Cooperative Movement (COP)	Any movement response that is clearly cooperative.
Morbid Content (MOR)	Content is characterized by death or damage, or is designated as being dysphoric.
Good Human Representation (GHR)	Positive representation of humans (i.e., Pure Human coding with +, o, or u Form Quality; see Exner, 2003, Table 9.1).
Poor Human Representation	Poor representations of humans (i.e., human responses that are Form Quality; see Exner, 2003, Table 9.1).
Personal (PER)	Reference to personal knowledge or experience is used to justify or clarify a response.
Color Projection (CP)	Identification of an achromatic portion of an inkblot as being colored.

* These Special Scores are rated as either Level 1, indicating a mild to modest level of cognitive slippage, or Level 2, indicating that the level of cognitive slippage is moderate to severe.

Core Section

The Core section includes the frequencies for the total number of responses (R), the total number for each of the determinants, and the following nine quantitative formulas:

1. **Lambda (L):**

$$\frac{\text{F (number of responses having only Pure F determinants)}}{\text{R} - \text{F (total R minus Pure Form answers)}}$$

   In calculating Lambda, only responses involving form are used (F, M, CF, and so on) and not determinants without form (C, C′, T, and so on).

2. **Experience Balance or Erlebnistypus (EB):** EB is the relationship between human movement responses and the weighted sum of the chromatic color responses. The ratio is expressed as Sum M: The Weighted Sum Color (WSumC). The Weighted Sum Color side of the ratio is calculated according to the following formula:

$$\text{WsumC} = (0.5) \times \text{FC} + (1.0) \times \text{CF} + (1.5) \times \text{C}$$

   All human movement responses are included in the formula, regardless of whether they are the major determinant of the response. Color naming responses are not included.

3. **Experience Actual (EA):**

$$\text{Sum of Human Movement} + \text{Weighted Sum Color}$$

4. **Experience Pervasive (EBPer):** Experience Pervasive is calculated by dividing the larger number in the EB ratio by the smaller one. This is done only when a marked difference (style) is evident in the EB ratio. This difference occurs when one of the following three criteria are met: (a) The value for EA is 4.0 or greater, (b) the value for Lambda is less than 1.0, and (c) "the value of EA falls between 4.0 and 10.0, one side if the EB must be at least two points greater than the other side. If the value of EA is more than 10.0, one side of the EB must be at least 2.5 points greater than the other" (Exner, 2003).

5. **Experience Base (eb):** The Experience Base ratio compares all nonhuman movement determinants (FM + m) with the sum of all the shading and achromatic color determinants. It is summarized by the following ratio:

$$\text{Sum FM} + \text{m: Sum C}' + \text{Sum T} + \text{Sum Y} + \text{Sum V}$$

6. **Experienced Stimulation (es):** This calculation merely requires adding together the two sides of the Experience Base ratio:

Sum of All Nonhuman Movement + Sum of All Shading or Achromatic Features

or

Sum FM + m + Sum C′ + Sum T + Sum Y + Sum V

7. **The D Score (D):** This is determined by first subtracting es from EA (EA – es) and designating whether the resulting number is a plus or minus number. The resulting raw score can then be converted into a standard score by consulting a conversion table provided in Exner (2003; Table 10.4).
8. **Adjusted es (Adj es):** All but 1m and 1Y (this also includes FY and YF) are subtracted from es.
9. **Adjusted D Score (Adj/D):** This is simply calculated by subtracting Adj es from EA (EA – Adj es). This produces a raw score that is converted to a standard score by using the same conversion table used in calculating the standard score for D score (see Exner, 2003; Table 10.4).

Ideation Section

This section consists of frequency data for M–, M, number of Level 2 responses, WSum6, and M with no FQ. In addition, there are three formulas:

1. **Active:Passive Ratio (a:p):** This is calculated by adding the total number of active movement responses and comparing it with the total number of passive movement responses:

$$M^a + FM^a + m^a : M^p + FM^p + m^p$$

2. **M Active:Passive Ratio (M^a:M^p):** In contrast to the previous active:passive ratio, this ratio refers only to active or passive responses relating to human movement and is calculated by simply inserting the total number of active human movements on the left side of the ratio and the total number of passive human movements on the right side.
3. **The Intellectualization Index:** This is calculated by multiplying the total number of Abstract (AB) responses by 2, and adding the sum of Art and Ay responses according to the following formula:

$$2AB + (Art + Ay)$$

The Affect Section

Rorschach indicators of affect include frequency of Pure C, S, and CP, as well as three quantitative formulas:

1. **Form-Color Ratio [(FC: (CF + C)]**: This ratio indicates the total number of form-dominated chromatic color responses, as compared with the absolute number of color-dominant chromatic responses. To calculate this formula, each of the chromatic color determinants is weighted equally as one. Cn determinants are also included on the left side of the ratio because they are considered color-dominant responses.
2. **Affective Ratio (Afr):** The Affective ratio is composed of the total number of responses to the last three cards, compared with those given to the first seven cards, or:

$$\frac{\text{Number of responses to Cards VIII + IX + } X}{\text{Number of responses to Cards I + II + III + IV + V + } VI \text{ + VII}}$$

3. **Blends: R (Complexity Index):** This compares the total number of blend responses (entered on the left side of the ratio) with the total number of responses (R).

Mediation Section

The Mediation section includes the total number of Popular responses and the total number of *S*– responses along with the following five percentages.

1. **Form Appropriate Extended (XA + %):**

$$\frac{\text{Sum of responses that have an FQ coding of +, o, or u}}{\text{R}}$$

2. **Form Appropriate—Common Areas (WDA%):**

$$\frac{\text{Sum of W + D responses with an FQ coding of +, o, or u}}{\text{Sum of W + D}}$$

3. **Distorted Form Quality (X – %):**

$$\frac{\text{Sum FQ} -}{\text{R}}$$

4. **Conventional Form (F + %):**

$$\frac{\text{Sum FQx + and o}}{\text{R}}$$

5. **Unusual Form (Xu%):** Unusual Form is a measure of the extent to which the contours of the inkblots have been used appropriately but unconventionally.

The Processing Section

This section includes three simple sets of frequency data—for Zf (total number of times an Organizational Activity response has occurred), PSV (Perseverations), DQ+ (Developmental Quality+), and DQv (Developmental Quality)—along with three ratios:

1. **Economy Index (W:D:Dd):** This index is developed by simply listing the total number of whole (W) responses on the left, the total number of D responses in the middle, and the total number of Dd responses on the right.
2. **Aspirational Ratio (W:M):** The ratio of W to M represents a comparison between the total number of whole responses (placed on the left side of the ratio) and the total number of human movement responses (placed on the right side).
3. **Processing Efficiency (Zd):** Processing Efficiency is a difference score. It is necessary to estimate what the Organizational Activity scores should be (Zest) by first summing the total number of times an Organizational Activity response occurred in a protocol (without taking into account the weightings). Next, the sum of all the weighted scores for Organizational Activity (ZSum) is calculated. Finally, Zest is subtracted from ZSum:

   Zsum – Zest

   This allows an estimate of how much more Organizational Activity was actually used compared to how much would have been expected to be used, based simply on the total number of Organizational Activity occurrences (without their weightings).

Interpersonal Section

This section is composed of three sets of frequency calculations—sum of Cooperative Movements (COP), Aggressive Movements (AG), and primary and secondary Food contents (Fd), sum of Pure H answers, the number of PER Special Scores, ratio of GHR:PHR Special Scores, Sum T (transcribed from the Core section), and a:p (transcribed from the Ideation section)—two formulas:

1. **Interpersonal Interest (Human Content) H:** The sum of responses that are Pure H is entered on the left side, and the sum of human interest contents—Hd and (Hd)—is entered on the right side.

   H + (H) + Hd + (Hd)

2. **Isolation Index (Isolate/R):** Calculation of this index requires noting the total number of content responses for Botany (Bt), Clouds (Cl), Geography (Ge),

Landscape (Ls), and Nature (Na). Contents for Clouds (Cl) and Nature (Na) are then multiplied by 2 and added to the number of responses for the rest of the contents. This sum is then divided by the total number of responses:

$$\frac{Bt + 2Cl + Ge + Ls + 2NA}{R}$$

The Self-Perception Section

This section includes five frequency tallies for:

1. Sum of Form-reflection and reflection-Form responses (Fr + rF).
2. Total number of Form Dimension (FD) responses.
3. Total number of responses that have morbid content (MOR).
4. Sum of all responses that have content related to Anatomy (An) or X ray (Xy; primary or secondary).
5. Sum of V (transcribed from Core section).
6. A sixth entry comprises the ratio composed of the number of Pure H contents and the sum of (H) + Hd + (Hd) on the right:

$$H: (H) + Hd + (Hd)$$

The final component of this subsection is a ratio related to the number of pair responses:

1. **Egocentricity Index [3r + (2)/R]:** This index gives three times the weighting to reflection responses (r) compared to pair responses (2) and compares these to the total number of responses (R):

$$\frac{3 \times (Fr + rF) + Sum(2)}{R}$$

Special Indices

Exner (1993) has developed the following six Special Indices:

1. Perceptual Thinking Index (PTI).
2. Depression Index (DEPI).
3. Coping Deficit Index (CDI).
4. Suicide Constellation (S-CON).
5. Hypervigilance Index (HVI).
6. Obsessive Style Index (OBS).

The procedure for calculating these indices is more complex than for the other formulas and is not covered in the present section. However, scoring criteria and cutoff

scores can be found in Exner (2003) and on the commercially available record form under a section designated as the *Constellations Worksheet.* Summaries of interpretive hypotheses for these indices are included in the next section of this chapter.

INTERPRETATION

The following description of interpretive information is meant to serve as a reference guide to alert Rorschach interpreters to a potentially wide range of possible interpretive hypotheses. Although the format is as concise as possible, interpreters should be aware of the tremendous variety inherent in most Rorschach data. Effective interpreters should also have this variety reflected in the wide number of possible interpretive hypotheses they generate. A mere labeling or simplistic "sign" approach should be avoided. Rather, clinicians must begin and end by continually being aware of the total overall configuration of the data. For example, the same number of C responses in two protocols can easily have quite different meanings, depending on the implications from, and interactions with, other aspects of the Rorschach data.

The typical sequence for Rorschach interpretation should follow the general conceptual model for testing developed by Maloney and Ward (1976) and discussed in Chapter 1 (see Figure 1.1 on p. 34). The model requires that clinicians initially take a propositional stance toward the protocol (phase 2). The purpose of this stage is to develop as many tentative hypotheses as possible, based on the quantitative data, verbalizations, and client history. The number and accuracy of these hypotheses depend on the individual richness of the data as well as on the individual skill and creativity of the clinician. The final stage is the integration of the hypotheses into a meaningful and accurate description of the person (phase 4). This involves rejecting, modifying, or confirming previously developed hypotheses (phase 3). When this has been accomplished, clinicians can integrate the Rorschach interpretations into the overall report itself (phases 5, 6, and 7).

In the description of different interpretive hypotheses, continual reference is made to "high" and "low" scores. These relative weightings are based on extensive adult normative data that have been accumulated on the Rorschach. For comparisons of scores on individual protocols with normative ratings, clinicians can refer to Appendix I on page 694, which provides means, standard deviations, and other relevant descriptive statistics for the different Rorschach factors and quantitative formulas. Clinicians interested in child and adolescent assessment (ages 6 to 16 years) can consult the much more extensive age-based norms for children found in Exner (2003); and Exner and Weiner (1995).

The sequence of presenting interpretive information is first organized around specific scoring categories (Location, Determinants, Contents, Special Scorings). These are followed by scorings according to the sections in the Structural summary. This begins with the Core section and then proceeds to the sections for Ideation, Affect, Mediation, Processing, Interpersonal, Self-Perception, and Special Indices. These later groupings should provide a conceptually consistent means of organizing relevant interpretive material around functional domains, thereby enabling the different interpretations to be more easily integrated into the psychological report. For example, if a practitioner is interested in understanding issues related to interpersonal relationships, he or she can note the Rorschach data relevant to this area of functioning. Similarly,

information related to dealing with affect can be noted in the section on affect. These interpretations can then be compared, contrasted, and modified with other assessment material on these dimensions. Table 10.8 outlines the different interpretive categories in the sequence in which they are presented for interpretation.

It should be noted that the Comprehensive System provides a slightly different organization of scores according to the following eight clusters (see Exner, 2003; Table 13.1):

Ideation
Affect Features
Cognitive Mediation
Information Processing
Interpersonal Perception
Self-Perception
Capacity for Controls and Stress Tolerance
Situation-Related Stress

Many of the domains in these clusters in Exner (2003) are quite similar to the organization of scores in the Structural summary presented in the previous section and again later in the following section on interpretation (i.e., both have scores relating to affect, mediation, processing, interpersonal, and self-perception). This somewhat different organization is partially because of space restrictions. In addition, the organization of the following section should also provide a quite clear introduction into the various scoring categories that should be easy to follow and also provide the reader with a fairly easy means of developing some working knowledge of the Rorschach.

The process of reading through the many interpretations in the remainder of the chapter can potentially be tedious because of the sheer quantity. To deal with the quantity of interpretations, it is recommended that the practitioner initially skim over the different sections and interpretations. Next, a Rorschach protocol might be obtained through actually administering and scoring a Rorschach, requesting one from a colleague, or using one from one of Exner's books. The practitioner can then go through each of the different categories and generate hypotheses based on the client's results. The hypotheses can be integrated into a description of the person, based on domains measured by the Rorschach variables. This sequence would optimally make the information relevant and engaging as well as enhance the development of actual clinical skills.

Location

In general, the area of the inkblot to which examinees choose to respond is a reflection of the overall style in which they approach their world. This is especially true for the manner in which they confront uncertainties and ambiguities in their lives. For example, one person might perceive only the most obvious and concrete aspects of a situation, whereas another might avoid important aspects of a stimulus by focusing on small details and neglecting potentially more significant issues. An analysis of

Table 10.8 Scoring and interpretive domains for the comprehensive system

Location

Whole Response (W)
Common Detail (D)
Unusual Detail (Dd)
Space (S)
Developmental Quality (DQ)

Determinants

Form (F)
Human Movement (M)
Animal Movement (FM)
Inanimate movement (m)
Color Chromatic (C)
Color Achromatic (C')
Shading—Texture (T)
Shading—Diffuse (Y)
Form Dimension (FD)
Pairs (2) and Reflections (Fr)
Organizational Activity (Z)

Content

Human and Human Detail (H, Hd)
Animal and Animal Detail (A, Ad)
Anatomy and X Ray (An, Xy)
Food (Fd)

Popular Responses

Special Scores

Deviant Verbalizations (DV)
Deviant Responses (DR)
Inappropriate Combinations (INCOM)
Fabulized Combination (FABCOM)
Contamination (CONTAM)
Inappropriate Logic (ALOG)
Perseveration (PSV)
Abstract Content (AB)
Aggressive Movement (AG)
Cooperative Movement (COP)
Morbid (MOR)
Good Human Representation (GHR)
Poor Human Representation (PHR)
Personal (PER)
Color Projection (CP)

Ratios, Percentages, Derivations

Core Section—frequency data (taken from previous sections includes total number of responses plus each of the frequencies of the determinants) and the following nine formulas:

1. Lambda (L)
2. Experience Balance or Erlebnistypus (EB)
3. Experience Actual (EA)
4. Experience pervasive (EBPer)
5. Experience Base (eb)
6. Experience Stimulation (ES)
7. D Score (D)
8. Adjusted es (Adj es)
9. Adjusted D score (Adj/D)

Ideation Section—frequency data for M–, M, number of Level 2 responses, WSum6, and M with no FQ. In addition

1. Active:Passive Ratio (a:p)
2. M Active:Passive Ratio (Ma:Mp)
3. The Intellectualization Index [2AB + (Art + Ay)]

Affect Section—frequency of Pure C, S. and CP and the following three formulas:

1. Form-Color Ratio [(FC:CF +C)]
2. Affective Ratio (Afr)
3. Complexity Index (Blend:R)

Mediation Section—number of Popular responses, the total number of S– responses, and the following percentages:

1. Form Appropriate Extended (XA+%)
2. Form Appropriate-Common Areas (WDA%)
3. Distorted Form (X–%)
4. Conventional Form
5. Unusual Form (Xu%)

Processing Section—frequency data for Organization Activity (Zf), Perseverations (PSV), Developmental Quality+ (DQ+), Developmental Quality-v (DQv), and three ratios:

1. Economy Index (W:D:Dd)
2. Aspirational Ratio (W:M)
3. Processing Efficiency (Zd)

Interpersonal Section—frequencies of Cooperative Movements (COP), Aggressive Movements (AG), Food Contents, sum of Pure H, number of Perseverations (PER), ratio of Good Human to Poor Human Representation (GHR:PHR), Sum T, and active:passive (a:p), and the following two formulas:

1. Interpersonal Interest (H + (H) + Hd + (Hd)
2. Isolation Index Bt + 2Cl + Ge + Ge + Ls + 2NA/R

Self-Perception Section—Sum Form-reflection and reflection-Form response, sum Form Dimension responses, sum morbid content, sum Anatomy, sum X ray, sum V, ratio of Pure H: (H) + Hd + (Hd), and the:

1. Egocentricity Index [3r + (2)/R]s

Special Indices

Perceptual Thinking Index (PTI)
Depression Index (DEPI)
Coping Deficit Index (CDI)
Suicide Constellation Index (S-CON)
Hypervigilance Index (HVI)
Obsessive Style Index (OBS)

Rorschach locations does not provide information regarding why people approach their world in a certain manner; rather, it is limited to a description of their particular style.

Rorschach locations can be divided into usual and unusual features, depending on the area of the inkblot that is used. Frequently used locations, if they are within the normal number and of good quality, usually reflect good ties with reality, intelligence, ambition, good reasoning, and an ability to generalize. Unusual locations involving rarely used areas of the blot are associated with neurotic symptomatology, such as fears, anxiety, and obsessive or compulsive tendencies. An extreme use of unusual features may reflect more serious psychopathology (Exner, 2003).

Whole Response (W)

The whole response is related to the degree to which a person can interact in an efficient and active manner with his or her environment. This is particularly true if the quality and organization of the responses are good. Whereas whole responses occur with the greatest frequency in children from 3 to 4 years of age (Exner & Weiner, 1995), there is a gradual decline in later childhood and adolescence until 30% to 40% of normal adult responses are wholes. The average adult ratio of whole:detail is approximately 1:2 (refer also to interpretation of W:M and W:D:Dd formulas).

High W Rorschach (1921/1941) originally believed that a high number of W responses reflected a person's ability to organize and integrate his or her environment. However, subsequent research has modified this belief. W responses do reflect intellectual activity, but this activity can be understood only by looking at the quality of W responses (relative number of W+) and the relative complexity of responses (Exner, 1993, 2003; Exner & Weiner, 1995). In considering the complexity of responses, it should be noted that W occurs with greatest frequency for Cards V, I, IV, and VI, and with lowest frequency for Cards X, IX, III, and VIII (S. J. Beck, 1945). Thus, W responses for the latter cards require significantly greater organizational activity. If good-quality responses and a high degree of organizing activity are both present, a high number of W responses would reflect good synthesizing and abstracting abilities (N. M. Smith, 1981), ambition (Schachtel, 1966), good ties with reality (Abrams, 1955; Levitt & Truumaa, 1972), and excellent problem-solving abilities (S. J. Beck, 1961; Rossi & Neuman, 1961).

Low W Low W responses can reflect depression (S. J. Beck, 1960; Rapaport, Gill, & Schafer, 1968) or anxiety (Eichler, 1951). If the frequency, quality, and complexity are low, more serious levels of maladjustment (Exner, 1974) are indicated, such as intellectual deterioration possibly related to brain damage (Goldfried, Stricker, & Weiner, 1971) or mental retardation (Allison & Blatt, 1964).

Common Detail (D)

Rorschach (1921/1941) originally conceptualized the D response as reflecting the degree to which a person perceives and reacts to the obvious aspects of a situation (Rorschach, 1921/1941). This is supported by more recent normative data in which adult nonpsychiatric groups and outpatients gave 62% and 67% of their responses, respectively, as D, whereas inpatient nonschizophrenics and inpatient schizophrenics gave 46% and 47% of their responses, respectively, as D (Exner, 1974). D tends to be

most frequent for Card X (Exner, 1993). Any interpretations relating to D should take into account the fact that a greater number of R is likely to increase the relative proportion of D when compared with other locations (also refer to the W:D:Dd ratio).

High D Often found in persons who overemphasize the concrete and obvious aspects of situations (S. J. Beck, 1961; Exner, 1993, 2003), high D requires less energy and less integration than making a W response. A high emphasis on D may further suggest that the person sacrifices the full use of his or her intellectual potential by merely focusing on the safe and obvious rather than probing into the more novel and unusual. This is sometimes reflected in the remitted schizophrenic who focuses on a relatively safe, conservative, and socially desirable response, which is suggested by pre- and post-treatment changes in D from 40% to 73%, respectively (Murillo & Exner, 1973).

Where D+ is high, an excellent level of developmental functioning and a concern with precision are likely (Goldfried et al., 1971). On the other hand, if D is high but the quality of responses is low, a severe level of maladjustment is indicated (Exner, 1974).

Low D Persons under stress show a decrease in D and a corresponding increase in Dd (Exner, 1974). Furthermore, low D can reflect inadequate perceptual habits (Klopfer, Ainsworth, Klopfer, & Holt, 1956), which may suggest brain damage (Reitan, 1955b). The proportion of D is lowest in young children and gradually increases with age (Ames, Metraux, Rodell, & Walker, 1974).

Unusual Detail (Dd)

The Dd response is considered to represent a retreat from a person's environment by focusing on details rather than either perceiving the whole situation or noticing the more obvious elements of the environment. A clinician would expect the number of Dd responses to comprise approximately 6% of the total R for a normal adult. However, Dd is frequently higher in the protocols of normal children and adolescents. For schizophrenics or severely impaired compulsives, the proportion of Dd can increase to 25% or more (Exner, 1974). When Dd is in good proportion to W and D, a healthy adjustment, in which a person combines initiative with an appropriate ability to withdraw, is reflected.

High Dd Persons with high Dd scores reflect a need to pull back from the ambiguities that may be contained in a whole response. When high Dd occurs in people with schizophrenia, it suggests an attempt to narrow their perceptions of their environment to make these perceptions more congruent with their inner world (Exner, 1986). If Dd perceptions are combined with movement, the hypothesis that the person's thought processes are impairing his or her perceptions is given further support (Exner, 1986).

Compulsives use Dd to focus on the details of a situation in an attempt to reduce their anxiety and exert more control over their perceptions. Their thought processes are not flexible enough to take in a sufficient number of whole responses. This rigidity becomes more exaggerated as the overall number of Dd responses increases and the size of each perception decreases.

Space (S)

A high number of S responses (three or more) is associated with negativism, difficulty in handling anger, and oppositional tendencies (Exner, 1993, 2003). However, in normal

populations, a moderate number of S responses probably does not relate to hostility (Martin, Pfaadt, & MaKinster, 1983) but may suggest some contrariness that is adaptive (striving for independence, constructive self-assertion). This is especially true if form quality is good (B. Klopfer et al., 1956; Z. Piotrowski, 1957). If S responses are high (three or more) and occur with poor form quality and/or poor primitive movements, a clinician should consider the presence of anger, hostility, and potential acting-out (Exner, 1993).

DQ+ and DQv

Developmental quality scores relate to a person's relative ability to analyze and synthesize information. A high DQ+ (above 9 or 10) is consistent with more intelligent, complex, and sophisticated persons. However, this greater complexity does not necessarily mean that the person is well adjusted or even that his or her cognitions are accurate (see Zd for an index of both efficiency and accuracy). A number of disorders are characterized by quite complex cognitive operations, yet they are not well adjusted. In contrast to DQ+, a higher proportion (three or more) of low Developmental Quality (DQv) responses indicates persons who are immature and less sophisticated (children, neuropsychologically impaired, intellectually disabled; Exner, 1993, 2003).

Determinants

Because the majority of research has been done on the determinants, they are frequently seen as the core of the Rorschach data. An analysis of a person's determinant score shows the psychological activity that he or she engaged in while the response was being created. It examines his or her unique styles of perception and thinking, and how these interact with one another. In general, research has isolated specific details of the determinants that could possibly lure the clinician into a rigid and potentially inaccurate "single sign" approach. Again, a Rorschach interpreter should focus on the interaction among a large number of variables to modify, confirm, or reject tentative hypotheses derived from any single determinant score.

Form (F)

The amount of pure F in a protocol has generally been used to indicate the extent to which the person can remove affect from a situation. The presence of form in a response represents a certain degree of respect for the standards of the environment and reflects intact reasoning abilities. It is seen both as related to attention and concentration and as an index of affective control or delay (Exner, 1993, 2003). This is reflected in the fact that inpatients with schizophrenia have a relatively higher percentage of Fu and F– responses than other groups. However, schizophrenics have increases in pure F following treatment (Exner, 1993), and a higher level of pure F for schizophrenics is associated with a better prognosis (Exner & Murillo, 1977). The presence of a pure F response does not necessarily mean that no conflict is present, but rather that the person is able to suspend temporarily the affect associated with a conflict. Conversely, people in emotional turmoil are likely to produce a significantly lower number of pure form responses, reflecting their inability to remove their affect from their experience. (See also interpretations for Lambda and the percentages in the Mediation section: X + %, F + %, X – %, S – %, Xu%.)

High Pure F Persons with a high pure F score either are highly defensive and constricted (Leavitt & Garron, 1982) or merely demonstrate a good ability to deliberately suspend or control their affect (S. J. Beck, 1945; B. Klopfer et al., 1956). When a person is in a more defensive position, the number of pure F responses increases. For example, pure F increases in populations of recovering schizophrenics (Goldman, 1960), perhaps as a result of their attempting to cautiously give a socially acceptable answer in which they have to limit their affect. Also, pure F is higher among paranoid schizophrenics than among other types of schizophrenics (Rapaport et al., 1968), reflecting their greater degree of organization and caution. Pure F also increases for persons who have been given some prior knowledge of the purpose of the test (E. Henry & Rotter, 1956) or who are requested to respond as quickly as possible (Hafner, 1958).

After ECT, pure F is usually higher (D. Kelly, Marguilies, & Barrera, 1941), which corresponds with patients' subjective reports of decreased affect. Also, alcoholics give more pure F responses than do psychopaths (Buhler & LeFever, 1947), and Leavitt and Garron (1982) have found an increase in F% in the protocols of patients having both psychological disturbances and low back pain.

Low Pure F If pure F is low, a person's level of turmoil is likely to be sufficiently high to prevent screening out his or her affective response to a situation. For example, acute schizophrenics who have difficulty reducing their level of affect have a low number of pure F responses (Exner & Murillo, 1973). Likewise, certain characterological disorders (Buhler & LeFever, 1947) and organic disorders, in which there is difficulty controlling impulses, both have a low number of pure F responses (Exner, 1974).

Human Movement (M)

Probably more research has been done on the M response than on any other Rorschach variable. Most of this research is consistent in viewing M as reflecting inner fantasies connected to the outside world. More specifically, M represents the bridging of inner resources with reality, or what might be described as "internalization of action" (Exner, 1993, 2003). M is also an inhibitor of outward behavior, even though that inhibition may be only temporary. A high proportion of M responses has been associated with creativity (Dudek, 1968) and introverted thinking (Kunce & Tamkin, 1981), and there is a close relationship between M and daydreaming (Dana, 1968; Page, 1957). Schulman (1953) has shown M's relation to abstract thinking in that a high number of M responses reflects both an active inner process and a delay in expressing behavior. Thus, M can be generally understood as involving deliberate inner experience. In its positive sense, M can indicate good ego functioning, ability to plan, impulse control, and ability to withstand frustration. In a more negative vein, it can suggest an overdeveloped fantasy life.

While interpreting M, it is important to look carefully at the different components of the response. For example, does the movement involve conflict or cooperation? A high number of aggressive movements has been shown to reflect a person who is generally more aggressive and who typically perceives relationships as characterized by aggressiveness (Exner, 1983). The degree of passivity in the movement is also likely to suggest that the person has more dependent and passive behaviors external to the test situation (Exner & Kazaoka, 1978). Specific interests might be projected into the

movement responses, such as the increased number of dance movements perceived in the protocols of physical education and dance students (Kincel & Murray, 1984). The clinician should also consider other data both from within the test and external to it. Further elaboration regarding M, especially as it relates to the person's degree of control of impulses, can be derived by referring to the EB and EA ratios.

High M High M responses, especially if they are M+, have been reported by some authors to be associated with high IQ (Abrams, 1955; Goldfried et al., 1971); other authors have not found this relationship (Mason & Exner, 1984). An alternative hypothesis is that high M is associated with increased creativity (Dudek, 1968; Richter & Winter, 1966). Dana (1968) has proposed that high M can represent any or all of several different psychological processes, including fantasy, an accurate sense of time, intellect, creativity, delay, and certain aspects of interpersonal relationships. Further studies include abstract thinking as an important correlate to M and to an introverted thinking orientation (Kunce & Tamkin, 1981). Because all of this involves motor inhibition, high M can also indicate a capacity to delay impulses (Pantle, Ebner, & Hynan, 1994).

A relatively high number of M responses suggests that the individual is overly invested in his or her fantasy life, which might be similar to a "Walter Mitty syndrome." With a high number of M– responses, the person is likely to be deficient in social skills and to have poorly developed interpersonal relationships (Molish, 1967; Weiner, 1966) or even psychotic symptoms (Phillips & Smith, 1953). H. Schmidt and Fonda (1954), for example, have found a high number of M responses in manic patients.

Low M In many respects, a low M response indicates the opposite of what is suggested by a high M. Persons, especially depressives, who have a difficult time using their inner resources, usually have low M scores (Ames, 1959; S. J. Beck, 1945; Exner, 1993). In addition, highly impulsive persons usually have a low number of M (Pantle et al., 1994). Demented elderly patients have been found to produce a low number of movement responses of all types (Insua & Stella, 1986). Low M is also associated with inflexible persons who have difficulty accepting and adjusting to change (Alcock, 1963; Goldfried et al., 1971; Rapaport et al., 1968). This inflexibility can, at least in part, be explained by a low level of empathy and a lack of imagination (Klopfer et al., 1956; Piotrowski, 1960, 1969a). Because successful psychotherapy involves both flexibility and a relatively active inner life, low M is indicative of a poor prognosis (Goldfried et al., 1971; B. Klopfer et al., 1956). Conversely, a high number of good-quality M responses are a positive prognostic indicator.

Animal Movement (FM)

Whereas human movement responses serve to mediate between the inner and outer environment, animal movement reflects more unrestrained emotional impulses in which there is less ego control. The impulses are more urgent, more conscious, and provoked by situations beyond the person's control. These observations are reflected in the higher number of FMs in children (Ames et al., 1971) and the aged (B. Klopfer et al., 1956), and they correlate positively with MMPI scales that measure irresponsibility, aggressiveness, and distractibility (G. Thompson, 1948). If persons are in situations in which they have little control, FM is likely to be increased. For example, FM has been found to

increase in persons experiencing physical restraint (Exner, 1979), in chronic amphetamine users (Exner, Zalis, & Schumacher, 1976), and among prostitutes who were addicted to drugs (Exner, Wylie, Leura, & Parrill, 1977). Human movement responses involve delay; animal movements do not. FM responses correspond to persons who complain of "racing thoughts" and have too much on their minds (Exner, 1986, 1993).

High FM A high number of FM responses suggests these persons are governed by their needs and urges. They generally have a difficult time delaying gratification and, therefore, rarely plan toward long-term goals (Exner, 1974). Typically, they are highly defensive and use intellectualization, rationalization, regression, and substitution as their primary means of reducing anxiety (Haan, 1964). If the FM responses are aggressive, it is more likely that they will be assaultive (Sommer & Sommer, 1958). The general, overall theme of high FM responses is that thoughts or feelings are occurring beyond the person's control. The number of FM responses for children (ages 8 to 16) is from 3.0 to 3.5, whereas adults have an average of approximately 3.5 (Exner, 1986).

Low FM Low FM is found among persons who are overly inhibited in expressing their emotions and may deny their basic needs (Exner, 1993; B. Klopfer & Davidson, 1962). For example, Ames et al. (1974) have associated low FM with a decreased energy level in children.

Inanimate Movement (m)

Similar to *FM,* the number of inanimate movement responses also provides an index of the extent to which persons are experiencing drives or life events that are beyond their ability to control. The drives reflected by m threaten people's adjustment in that they are helpless to effectively deal with them (B. Klopfer et al., 1956). This helplessness is usually related to interpersonal activities (Hertz, 1976; B. Klopfer et al., 1956; Z. Piotrowski, 1957, 1960). For example, Exner (1993) has found one or more m responses in the records of both inpatient and outpatient schizophrenics, and Piotrowski and Schreiber (1952) found no m scores in the records of successfully treated patients. The number of m responses is also more frequent with juvenile delinquents, to the extent that, by 16 years of age, they perceive an average of one per protocol (Majumber & Roy, 1962). The view that m represents threat from the external world is supported by the observation that sailors at sea produced significantly more m during a severe storm (Shalit, 1965). This is also consistent with the finding that normal subjects exposed to uncontrollable laboratory-induced stress (McCown, Fink, Galina, & Johnson, 1992) and those given amphetamines (W. Perry et al., 1995) had temporary increases in m. Similarly, paratroop trainees had an increase in m just before their first jump (Armbuster, Miller, & Exner, 1974) as did elective-surgery patients just before surgery (Exner, 1993; see also the interpretation of experience base [eb]).

High m The presence of m should serve as a warning sign to indicate a marked presence of conflict and tension. Subjects probably see themselves as surrounded by threatening persons and are unable to reconcile themselves with their environment. A related finding by C. Thomas and Duszynski (1985) is that the word *whirl* (or similar words) was found more frequently in the protocols of persons who later committed suicide.

Although the use of these "whirl all" words may not have necessarily been formally scored as FM or m, there are clear similarities between these classes of responses. To gain a more complete understanding of the individual meaning of m, clinicians should investigate the possible resources and the characteristic means of resolving conflict by looking at M, sum C, frequency of D and S, and the accuracy of their perceptions as reflected in F + % and X + %.

Color Chromatic (C, CF, FC, Cn)

The manner in which color is handled reflects the style in which a subject deals with his or her emotions. If color dominates (C, CF, Cn), affect is likely to be poorly controlled and disorganized. In such cases, affect is disruptive and the person could be expected to be emotional, labile, and overreactive. If the responses are more dominated by form (FC), affect will be more delayed, controlled, and organized. For example, it has been demonstrated that subjects who could effectively delay their responses in a problem-solving task had a higher number of FC responses in their protocols, whereas those who had difficulty delaying their responses had more CF and C responses (H. Gill, 1966; Pantle et al., 1994). They also found that a positive correlation exists between individuals having color-dominated responses and measurements of impulsiveness. However, if the number of color-dominated responses is used to determine impulsiveness, the implications of D scores, form quality, number of Y responses, and relative number of color-dominated responses (FC:CF + C) should also be taken into account. Furthermore, the chromatic cards produce a greater frequency of aggressive, passive, and undesirable contents than do the achromatic cards (Crumpton, 1956).

Adult nonpatients have between 1.5 to 2.5 times more form-dominated color than color-dominated responses [FC/(CF + C)]. This contrasts with the average patient group, which generally has an equal number of FC to CF + C responses (Exner, 1993). Pure C responses are also predominant in the protocols of very young children, as is color naming (Cn; Ames et al., 1974; Exner, 1986). (See also interpretation of the FC:CF + C formula.)

High C and Cn Individuals with a high proportion of color-dominated responses typically have little regard for the adaptiveness of their expressions, and they discharge their emotions in an impulsive manner (H. Gill, 1966; Pantle et al., 1994). This suggests that higher cognitive abilities have been suspended or possibly overwhelmed by affective impulses. Stormant and Finney (1953) were able to differentiate between assaultive and nonassaultive patients based on the assaultive patients' having a higher number of poor-quality color responses. Likewise, Townsend (1967) found a higher level of aggressiveness in adolescents who produced a greater-than-average number of CF responses combined with an absence of human movement. In general, a high number of color-dominant responses suggests that the person is more labile, suggestible (Linton, 1954; Mann, 1956; Steisel, 1952), sensitive, and irritable (R. Allen, 1954; D. A. Shapiro, 1960), and has difficulty delaying his or her responses during problem-solving tasks (Pantle et al., 1994).

Color naming suggests that the person is giving a concrete response to the stimuli, and the response is primitive and poorly conceptualized. Although research is inconclusive, color naming typically seems to occur in severe disorders for adults, such as

organic impairment. This is somewhat supported in that some brain-damaged subjects show an increased interest in color and seem to be more "stimulus bound" in their perception of it (see Lezak, 1995; Stuss, Gow, & Hetherington, 1992). Color naming is not unusual in the protocols of young children, particularly if intellectually disabled (Exner & Weiner, 1995).

Low C and CF A total absence of (or at least a very low frequency of) C and CF occurs more frequently with depressed or psychosomatic patients and those with a low level of spontaneity who consistently dampen and overcontrol their emotional expression (C. Costello, 1958; Exner, 1993, 2003). If other suicidal indicators are present, a low color-dominant protocol may give additional support to the presence of suicidal tendencies (Goldfried et al., 1971). Low C and CF responses from schizophrenics can be a good sign for successful treatment (Stotsky, 1952).

High FC A moderately high number of FC responses can indicate a good level of integration between controlling emotions and appropriately expressing them (S. J. Beck, 1945; B. Klopfer et al., 1956). Typically, this level of FC responses indicates that individuals have the ability to develop good rapport with others (Allison, Blatt, & Zimet, 1968), low levels of anxiety (Greenwald, 1991), and a capacity to learn under stress (Phillips & Smith, 1953). The prognosis for therapy is good (Goldfried et al., 1971) because they can experience emotions yet also conceptualize and give form to the expression of these emotions. Beck has stated that a moderately high number of FC responses indicates that schizophrenia is unlikely. In children, it may reflect the effects of overtraining with a corresponding decrease in natural spontaneity (B. Klopfer et al., 1956; D. A. Shapiro, 1960). In adult populations, it may also reflect overcompliance and a dependent personality (Schafer, 1954).

Low FC Low FC suggests poor emotional control (B. Klopfer et al., 1956), which is likely to negatively affect interpersonal relationships (Z. Piotrowski, 1957). Low FC can also indicate anxiety states (Rapaport et al., 1968) and gives support to a hypothesis of schizophrenia if other indicators of schizophrenia such as poor-quality responses are present (S. J. Beck, 1945; Thiesen, 1952).

Color Achromatic (C′, C′F, FC′)

Achromatic color responses constitute one of the least researched areas of the Rorschach. However, it has been suggested that C′ responses reflect constrained, internal, and painful affects. In other words, there is a dampened emotional expressiveness in which the person is cautious and defensive. Exner (1993) has referred to C′ as the psychological equivalent of "biting one's tongue, whereby emotion is internalized and consequently creates some irritation" (p. 386). Thus, it relates not only to painful emotions, but also to affective constraint and defensiveness. Most Rorschach systematizers have consistently used C′ as an index of depression. In considering the meaning of achromatic color responses, a clinician should look at the relative influence of form. If form is dominant (FC′, there is likely to be definition and organization to the affect, with a stronger ability to delay the behavior. On the other hand, dominant C′ responses suggest the immediate presence of painful emotions.

The average number of achromatic color responses for nonpatients is 1.49 (Exner, 2003). In contrast, depressives have an average of 2.16 and .83 for character disorders (Exner, 1993).

High C′ C′ occurs most frequently among patients who constrain their emotions, such as psychosomatics, obsessive-compulsives, and depressives (Exner, 1991, 1993, 2003). The pain and constraint associated with these emotions may adversely affect these persons' overall level of adjustment. An absence of shading responses combined with a large proportion of C′ responses has been suggested as predictive of suicidal gestures (Exner, 1974).

Shading—Texture (T, TF, FT)

Texture responses represent painful emotional experiences combined with needs for supportive interpersonal relationships (S. J. Beck, 1945, 1968; B. Klopfer et al., 1956). For example, recently divorced or separated subjects averaged 3.57 texture responses per protocol (SD = 1.21) as compared with 1.31 for matched controls (SD = 0.96; Exner & Bryant, 1974). Persons with a high number of texture responses reach out, although they do so in a guarded and cautious manner (Hertz, 1976). If form plays a relatively insignificant role and texture is predominant, subjects tend to feel overwhelmed by painful experiences, which would probably be sufficiently intense to disrupt their ability to adapt. Conversely, if form dominates (FT), not only is the pain likely to be more controlled, but also the need for supportive contact from others would be of primary concern (S. J. Beck, 1968; B. Klopfer et al., 1956). Coan (1956) has suggested that a combination of movement and texture responses relates to inner sensitivity and empathy. If chromatic color and texture occur together, the subjects' behaviors would not only be less mature in seeking affection, but also more direct and unconstrained (Exner, 1974).

Responses in which texture dominates show an increase through childhood, reach a maximum by 15 years of age, and gradually subside over the next few years until a form-dominated texture response is most characteristic in late adolescence and adulthood (Ames et al., 1971). Nonpatient populations average one texture response per record, whereas psychiatric populations average two or more per record (Exner, 1993, 2003). They usually appear 10 times more frequently on Cards IV and VI than on the other cards (Exner, 1993).

High T or TF High scorers for T or TF are characterized as having intense needs for affection and dependency (Exner, 1993; Greenwald, 1991). Oversensitivity in personal relationships may result, to the extent that they may have a difficult time in reconciling the intensity of these needs with what they can realistically expect from their relationships. They are open to their environment, but they approach it with a cautious sensitivity.

Low T The absence of any T responses may suggest an emotional "impoverishment" in which the person has ceased to look for meaningful emotional relationships (Exner, 1993, 2003). For example, inpatient depressives have the lowest average number of texture responses but the highest number of diffuse shading responses (Y; Exner, 1974).

Likewise, psychosomatic patients give fewer T responses than other types of patients (F. Brown, 1953), which would correspond with their constrained expression of affect. In general, T, Y, and C′ all represent an "irritating emotional experience" such as anxiety, tension, apprehension, and internal discomfort (Exner, 1993).

Shading—Dimension (Vista; V, VF, FV)

Rorschach systematizers have generally considered Vista responses, especially pure V, to represent a painful process of self-examination in which the person creates a sense of distance from self to introspect (Exner, 1993). This introspection usually involves depression and a sense of inferiority. However, if the V responses are dominated by form, introspection is still suggested but the process is unlikely to be emotionally painful. This is in contrast to the negative type of self-examination associated with pure V. Even a single pure V response in a Rorschach protocol can be an important indicator.

In normal populations, V responses occur, on average, 0.28 per record. Depressed inpatients average 1.09, and schizophrenics and character-disordered persons average 0.60 and 0.24, respectively (Exner, 1993, 2003). It is extremely rare for V to appear in the protocols of children, but it occurs at about the same rate among adolescents as it does for adults (Exner, 1993; Exner & Weiner, 1995).

High V Pure V responses created by depressed patients indicate a deep level of self-critical introspection (Exner, 1993). Stutterers also produce more pure V responses (Light & Amick, 1956), as do alcoholics (Buhler & LeFever, 1947), reflecting the painful self-criticism that usually occurs in these patient groups. V responses have also been suggested as an index of suicidal risk and are an important part of Exner's Suicidal Constellation (Exner, 1991, 2003). Although shading (and combined color and shading) responses in themselves are probably ineffective in discriminating successful from nonsuccessful attempters, these responses may suggest a more stable suicidal trait (Hansell, Lerner, Milden, & Ludolph, 1988). However, Exner's Suicidal Constellation, composed of 12 possible signs (high number of morbid responses, es greater than EA, and so on) with a cutoff of 8 or more, has been able to effectively discriminate persons who are serious suicidal risks (Exner, 2003).

Low V The absence of V is usually a positive sign, and the presence of a single form-dominated V merely represents the ability to introspect (Exner, 2003). Although a certain degree of pain may be involved with the introspection, the more important fact is that the resulting information can be integrated and eventually used productively.

Shading—Diffuse (Y, YF, FY)

B. Klopfer et al. (1956) and S. J. Beck (1945) have described Y as representing a sense of helplessness and withdrawal, which is frequently accompanied by anxiety and is often a response to ambiguity. Beck further elaborated that subjects with a high number of Y responses are experiencing psychological pain and have resigned themselves to their situation. The same general rule for looking at the influences of form (F) in relation to Vista (V), texture (T), and color (C, C′) also applies for shading-diffuse. When *F* is dominant, subjects are more able to delay their behavior, and their experience is more controlled, organized, and integrated. This ability to delay behavior also gives them time to mobilize their resources. When Y is dominant, there is a much greater

sense of being overwhelmed. Although these individuals are characteristically withdrawn, any expression of pain and helplessness is direct. Because there is little ability to delay their impulses, they do not have enough time to mobilize their resources.

In the general population, 86% of people give at least one Y (Y, YF, or FY) response. Schizophrenics give more Y responses than nonpatients and outpatients, and nonschizophrenic patients give twice the number of Y responses than normals do (Exner, 1993, 2003). To accurately understand the meaning of Y responses, the clinician should look for other indicators of coping. In particular, these might include the number and manner in which pure form is used, the quality of organization, and the number of human movement responses. If there is a high number of Y and these "coping indicators" are absent, the person is likely to be overwhelmed and probably unable to adapt or respond effectively (Exner, 1993).

High Y A high number of Y is associated with anxiety (S. J. Beck, 1961; B. Klopfer & Davidson, 1962) and a constrained expression of emotions, even though the experience of these emotions may be direct (Salmon, Arnold, & Collyer, 1972). It is more frequent in the protocols of depressed patients and outpatients (Exner, 1978). High Y is also associated with a sense of resignation to life events and an attempt to create distance between oneself and the environment (Elstein, 1965). Y is higher in alcoholics (Buhler & LeFever, 1947) and increases during stress, such as before examinations (Ridgeway & Exner, 1980), surgery (Exner, Thomas, Cohen, Ridgeway, & Cooper, 1981), uncontrollable laboratory-induced stress (McCown et al., 1992), and situational crises (Exner, 1993, 2003). Because m and Y assess similar constructs, they should be considered together.

Low Y Ambiguity is purposefully built into the test situation. Some Y, usually FY, can, therefore, be expected to occur in any protocol. Exner's (1993, 2003) normative groups of adult nonpatients had an average of 0.57 Y responses (SD = 1.00), compared with 2.12 for schizophrenics (S = 52.62), and 1.81 for depressives (SD = 1.40). The total absence of Y suggests an extremely indifferent attitude toward ambiguity (Exner, 1993)

Form Dimension

Form Dimension (FD) is unique to the Comprehensive System and was included because it seemed to be both an empirically distinct category and a source of some interpretive significance. The research that exists suggests that a high number of FD responses are related to introspection and self-awareness. For example, a relatively high number of FD responses have been found among persons who are introverted and are involved in the later phases of insight-oriented therapy, and among patients who have completed a wide number of other forms of therapy (Exner, 1993). FD responses occur more frequently among nonpatients (M = 1.18) than among other patient groups—including schizophrenics (M = 0.60) and depressives (M = 0.82), and is particularly low among character disorders (M = 0.33; Exner, 1993, 2003).

High FD High FD suggests that the person is introspective, self-aware, and able to delay and internalize behaviors. This introspection is not likely to be painful unless high FD is also associated with other indicators, particularly V, which would then suggest affective difficulties.

Pairs (2) and Reflections (rF and Fr)

Research on pairs and reflections has been linked both conceptually and empirically to relevant aspects of personality. Self-absorption is suggested if either of these categories is high (Exner, 1991, 1993, 2003). However, this does not necessarily mean that the individual is pathological. For example, a high number of reflections were found among nonpatients in occupations that encourage a high level of self-worth, such as performing artists and surgeons (Exner, 1993). Whereas reflections occurred in only 7% of adult outpatients, they occurred in a full 20% of the protocols of character disorders and 75% of the records of antisocial groups (Exner, 1993). It is fairly common for children between the ages of 5 and 10 to have a high number of reflection (and pair) responses, but they usually decrease by adolescence, when individuals move to a less egocentric style of functioning (Exner, 1993; Exner & Weiner, 1995).

High Pairs (2) and Reflections (rF and Fr) A high number of pairs and reflections suggests these persons are self-absorbed and have an inflated sense of self-worth. They might have an exaggerated sense of self-pride, with strong strivings toward status. If they are unable to achieve affirmation from their environment, they might become depressed and negativistic. This could potentially lead to difficulties in maintaining deep and meaningful relationships (Exner, 1991, 1993, 2003).

Organizational Activity (Z)

The relative extent to which a person efficiently and effectively organizes the disparate aspects of the inkblots will be reflected in the scoring for Organizational Activity. The possibility that Organizational Activity is conceptually related to intelligence is given some empirical support in that moderate correlations (.54) have been found with the Wechsler-Bellevue Full Scale IQ and an even higher correlation of .61 exists with the Wechsler Vocabulary subtest (see Exner, 1993). Adults and younger nonpatients will have frequencies of Organizational Activity ranging between 9 and 13 (Exner, 1993, 1995). Among psychiatric patients, lower organizational activity has been noted among depressed patients (Hertz, 1948). In contrast, quite high levels of organizational activity have been found among patients who projected organized delusions (S. J. Beck, 1945; see also the interpretation for Processing Efficiency in the Processing Section topic later in the chapter).

High Zf (>13) High scores indicate that the degree of effort expended to process information is more extensive than required or expected. Thus, these persons may have a high level of intellectual striving in which they carefully and precisely work with their perceptions (Exner, 1993).

Low Zf (<9) With scores of less than 9, it would be expected that the client expends less effort than needed or required to adequately process information.

Content

The different content categories are generally considered to contain information relating to a person's needs, interests, preoccupations, and social interactions. Positive

correlations have been found between a large variety of contents and intelligence (Exner, 1986). Research has also shown that, whereas a high variety of contents is associated with intellectual flexibility, a low variety suggests intellectual constriction and rigidity. Persons' occupational interests are often represented in a higher number of contents relating to their specific career choices. For example, biologists and medical personnel usually give a higher number of anatomy responses than the general population (Exner, 1974). This may indicate that these persons merely have an interest in their career, but it could suggest they are overconcerned with their career to the extent that they neglect other areas of their lives, perhaps even impairing their overall level of adjustment. For example, biologists who see only nature contents may be using a preoccupation with their careers to withdraw from interpersonal relationships (Exner, 1974).

When interpreting Rorschach content, it is important to look at the variety of contents, the number of each content, and their overall configuration, as well as the implications other Rorschach factors may have for the meaning of the content scorings. It is usually essential to consider the age of the subject and to use age-appropriate norms. For example, children usually have significantly fewer human and human detail responses than adults, and the variety of their contents is also lower (Ames et al., 1974; Exner & Weiner, 1995). Another important step is to study contents relating to aggressiveness (fire, explosions, etc.), facial features, and orality. Although the focus of the Comprehensive System is on a quantitative approach to the Rorschach, symbolical considerations can also be extremely important in conducting a more qualitative analysis. The following section provides general information on the meaning associated with human and animal contents. Further interpretive material can be found in the interpretation of quantitative formulas relating to contents. See Intellectualization Index, Isolation Index, Interpersonal Index, (H) + (Hd) : (A) + (Ad), and H + A : Hd + Ad.

Human and Human Detail [H, Hd, (H), (Hd)]

Human responses constitute one of the most thoroughly researched contents. S. J. Beck (1961), in general agreement with other researchers, has found that H and Hd gradually increase with age until the median for 10-year-old children is from 16% to 18%. This remains unchanged through adolescence, and the overall adult proportion of 17% is eventually reached. Exner (1974) found that, whereas adult nonpatient H + Hd responses were 19% of the total adult outpatients' responses, schizophrenics had a lower total of 13%. He also demonstrated that the ratio of human to human detail (H : Hd) for nonpatients was 3 : 1. In contrast, schizophrenics' average ratio was approximately 1 : 1 and outpatients' ratio was 2 : 1. Molish (1967) suggests that when there is an increase in Hd compared with H, the subject is prone to use constricted defenses. Others have theorized that the increase suggests intellectualization, compulsiveness, and a preoccupation with the self that restricts the degree of contact with others (B. Klopfer & Davidson, 1962). S. J. Beck (1945) associated high Hd with anxiety, depression, and a low intellectual level (see also the quantitative formula for Interpersonal Index).

High H A high number of human contents occurs among individuals who have a wide-ranging interest in people (S. J. Beck, 1968), are more likely to have high self-esteem (S. Fisher, 1962), and possess greater intelligence (S. J. Beck, 1968; Rawls & Slack, 1968). A higher H content has been consistently found to be associated with a greater

likelihood of successful psychotherapeutic treatment (Goldfried et al., 1971; Weiner & Exner, 1991). As might be expected, human responses are more frequent in the records of psychologists and anthropologists (Roe, 1952).

Low H An unusually low number of H contents suggests a low level of empathy and a withdrawal from interpersonal relationships (Allison et al., 1968; T. Kahn & Giffen, 1960). The overall H% has been found to be lower for schizophrenics than for normals (Exner, 1993). The prognosis for successful psychotherapy with low H scorers is poor (Z. Piotrowski & Bricklin, 1961), and if the low score is accompanied by a low number of M responses, their termination from therapy is likely to be abrupt, probably because a high level of anxiety is combined with a low intellectual level (Affleck & Mednick, 1959; Rapaport et al., 1968). The complete absence of human content is unusual and, with the exception of very young children, it is likely to indicate psychopathology characterized by difficulties with identity (Exner, 1978).

Animal and Animal Detail (A and Ad)

Most of the literature indicates that animal content is associated with the obvious aspects of adaptiveness and the most concrete features of reality testing (Draguns, Haley, & Phillips, 1967). Because animal contents are the easiest to perceive, their presence suggests that examinees are using routine and predictable ways of responding. Conversely, a low number of animal responses suggests highly individualistic persons who see their world in their own personal and unique ways.

Animal responses occur more frequently than any other content category and comprise 38% to 48% of the normal adult record (Cass & McReynolds, 1951) with a slightly higher amount for children (S. J. Beck, 1961). Schizophrenics and outpatients average 31% and 41%, respectively, whereas depressives score much higher, averaging 41% per protocol (Exner, 1974). Other studies have found that the percentage of A responses is low for manics (Kühn, 1963; H. Schmidt & Fonda, 1954) and high for alcoholics (Buhler & LeFever, 1947).

High A High A suggests a predictable, stereotyped manner of approaching the world (B. Klopfer & Davidson, 1962; Levitt & Truumaa, 1972)—a manner often associated with depression and the use of constrictive and conforming defenses (S. J. Beck, 1945, 1960). There has been some evidence to suggest that high A responses are a sign of brain damage (Goldfried et al., 1971), but variables from without as well as within the test should be carefully considered before making this diagnosis.

Low A Persons who are spontaneous, nonconforming, unpredictable, and of higher intelligence often have a low number of A responses (R. Allen, 1954; T. Kahn & Giffen, 1960).

Anatomy (An) and X Ray (Xy)—Body Concern

Because An and Xy both measure concern with the body, they are considered together. Anatomy (An) responses have been well researched, and, along with human and animal contents, anatomy is one of the most frequently occurring responses (average of 0.6 for nonpatient adults). Anatomy content has an obvious connection with concern for the

body, and the literature supports this connection in that it occurs more frequently for persons preparing to undergo elective surgery (Exner, Armbuster, Walker, & Cooper, 1975) and among psychosomatic patients (Shatin, 1952). Anatomy responses also occur with greater frequency with the onset of psychological difficulties related to pregnancy (Zolliker, 1943). As might be expected, Anatomy responses occur more often in the protocols of biologists and persons with medical training (Roe, 1952; Schachtel, 1966). A review of the literature by Draguns et al. (1967) concluded that anatomy content can serve as an index of the degree of involvement persons have in their inner fantasy life or may reflect physical changes such as illness, puberty, or pregnancy. Exner (1993) has also suggested that anatomy content is associated with withdrawal from the environment and obsessive defenses.

The relative proportion of anatomy to Xy responses is an important consideration. Although anatomy responses are generally low for both psychiatric and nonpsychiatric groups, a combined anatomy and Xy score allows for a clearer differentiation between the two groups. Whereas the combined An and Xy responses for a nonpsychiatric group give an average of only 0.6 responses, outpatients give 1.5, schizophrenics give 1.4, and nonschizophrenic patients give 1.8 (which accounts for 9% of this last group's total number of responses; Exner, 1974, 2003). Xy responses have been found to be particularly high for schizophrenics with bodily delusions (average of 2.2) and depressed patients with concerns related to bodily functioning (1.7; Exner, Murillo, & Sternklar, 1979). Anatomy responses occur most frequently for Cards VIII and IX, and Xy responses are most frequent for Card I. Exner (1974) suggests that Xy responses reflect a concern with the self that is painful, but that these subjects are attempting to deal with this pain by distancing themselves from it or at least disguising their responses to it. On the other hand, anatomy responses reflect a process in which the person focuses more directly on the stress and there is a more direct emotional release.

A high number of An and Xy responses are associated with hypochondriasis (Carnes & Bates, 1971; E. Wagner, 1973) and psychosomatic conditions (Shatin, 1952). In congruence with these disorders, there are likely to be intellectualizing defenses (Allison et al., 1968), anxiety (E. Wagner, 1961), obsessive traits, and withdrawal (Exner, 1974). High An may also reflect a concern with physical functioning because of aging (Ames, Metraux, Rodell, & Walker, 1973) or career choice (Roe, 1952). Schizophrenic patients sometimes give an unusually high number (eight or more) of anatomy responses (Brar, 1970; Goldfried et al., 1971).

Food (Fd)

A high number (two or more) of food contents (primary or secondary) suggests dependency. High scorers would be expected to request extensive help and guidance from others, have difficulty making independent decisions, and be naive in their expectations of others (Exner, 1993).

Popular Responses

The number of Popular responses reflects the subjects' degree of similarity to most people, the extent to which they conform to social standards, and the relative ease with which they can be influenced in interpersonal relationships. Persons who reject conventional modes of thinking give a significantly lower number of Populars than those who

are conforming and relatively conventional. With Exner's (1993) scoring system (see Table 10.6), the average number of P responses for nonpsychiatric subjects is 6.9 (SD = 1.39). Outpatients and nonschizophrenic patients, likewise, give approximately 7 per record, whereas inpatient schizophrenics give 4 or less, characterologically disordered persons give approximately 5, and depressives have slightly more than 5 (Exner, 1993).

High P High P suggests that the subject is experiencing anxiety related to a fear of making mistakes and, therefore, clings to common perceptions as a way to achieve approval. These individuals can be described as conventional, overconforming, guarded, and, frequently, depressed (Exner, 1974; Levitt & Truumaa, 1972; Weiner, 1961).

Low P The lowest number of P responses is given by inpatient schizophrenics, which is consistent with their poor contact with reality. They can be described as poorly adjusted, detached, aloof from their environment, and unable to see the world as others see it. Molish (1967) has suggested that if neurotic subjects, especially obsessive-compulsives, have low P, the possibility of latent schizophrenia should be investigated. Patients diagnosed as having character disorders also have low P, which reflects their rejection of conventionality and their lack of conformity.

Because Populars are extremely common for Cards I, III, V, and VIII, an absence of them from these cards is significant in that it more strongly suggests the trends just discussed. However, the assumption that low P responses alone confirm maladjustment should be approached with caution. Low P subjects who have good form quality (F + % and X + %) and whose organizational activity is also good are likely to be creative individuals who are avoiding common or ordinary perceptions and want to extend their imagination. If organization and form quality are poor, there is a high likelihood that the psychopathological dimensions are more predominant.

Special Scores

Deviant Verbalization (DV), Deviant Responses (DR), Incongrous Combination (INCOM), Fabulized Combination (FABCOM),Contamination (CONTAM), Inappropriate Logic (ALOG)

The first six of the Special Scores were included in the Comprehensive System to detect the presence of cognitive slippage. Illogical, dissociated, fluid, or circumstantial thinking is particularly likely if there are any Level 2 scorings for the first four scores (Exner, 1991). This is consistent with the finding that virtually no Level 2 DV or DR responses occurred among nonpatients but fully 1.90 Level 2 DRs have been noted among schizophrenics (Exner, 1993). However, there is no specific interpretation for each of the six categories. Instead, they are used collectively to detect the presence and seriousness of cognitive distortions. The relative seriousness is indicated in part by the type of Special Score. Mild distortions are suggested by the presence of scores for DV (Level 1), INCOM (Level 1), or DR (Level 2), and moderate distortions are suggested by the presence of DV (Level 2), FABCOM (Level 1), INCOM (Level 2), and ALOG. The most serious degree of cognitive distortion is suggested if patients have Special Scores for DR (Level 2), FABCOM (Level 2), and CONTAM.

A further means of analyzing the first six Special Scores is by noting the relative elevation of WSUM6, which is simply a sum of the different weightings given to the Special Scores (see the section on "Scoring"). The WSUM6 for nonpatients is 3.2, indicating that normals do include at least some of the Special Score responses. In striking contrast are schizophrenics, who have a WSUM6 of nearly 45 (Exner, 1993). However, the presence of Special Scores does occur among children under 10 but gradually decreases during adolescence (Exner & Weiner, 1995). The general interpretation for the first six high Special Scores is that there is cognitive distortion. The interpretive hypothesis, especially with a high WSUM6, is that there is a serious disregard for reality, strained reasoning, faulty cause-and-effect relationships, loose associations, disorganized thinking, and poor ability to focus on tasks (Exner, 1991, 1993). This ability of the Rorschach to detect the bizarre and illogical processes of schizophrenia is probably one of its best validated features (Vincent & Harman, 1991), and there is some evidence that it is sensitive to these changes in thought processes even before their clinical manifestation (G. Frank, 1990).

Perseveration (PSV)

The presence of perseveration suggests some difficulty in cognitive shifting. Thus, the individual may have either a permanent or a temporary difficulty with rigidity or inflexibility in information processing or decision making (Exner, 1993).

Abstract Content (AB)

The presence of one or more abstractions suggests intellectualizing defenses (see Intellectualization Index).

Aggressive (AG) and Cooperative Movement (COP)

It is useful to consider AG and COP together. If there is an absence of scores on either category, it suggests that the individual is aloof, somewhat uncomfortable in social situations, and on the periphery of group situations. In contrast, if COP is high (two or more) and AG is low (zero or one), the person is likely to be perceived by others as trustworthy, cooperative, and easy to be around (Exner, 1993). It is also a favorable prognosis for psychotherapy. If COP is low (less than three or especially zero) and AG is high (greater than two), the person's interactions are likely to be forceful or even aggressive and hostile (Exner, 1993). Given these interpretations, it might be speculated that high scores on both COP and AG would suggest some conflict regarding the appropriate and preferred mode of responding and would result in inconsistent interpersonal behaviors (i.e., passive aggressive interactions).

Morbid Content (MOR)

Although the presence of one MOR is not unusual in the records of nonpatients, two or more suggest pessimism, a negative self-image, and possible depression and is consistent with a diagnosis of PTSD (Weiner, 1996). If three or more MOR responses are present, it is both a strong indicator of depression and one of several indicators for suicide risk (see the Suicide Constellation; Exner, 1991, 1993). MOR content is likely to have unique meaning for the person and can often be interpreted symbolically and qualitatively.

Good Human Representation (GHR) and Poor Human Representation (PHR)

GHR and PHR are considered as dichotomous categories. Persons with a high number of GHRs are usually highly regarded by others, well adapted, competent, and are reasonably free from chaos (Exner, 2003). In contrast, persons with psychiatric histories typically give a low number of GHRs. If they also give a high number of PHRs, they are also likely to report histories of interpersonal difficulties, are socially inept, and are interpersonally ineffective (Exner, 2003).

Personal (PER)

Scores of three or more suggest a defensive authoritarian stance in which the individual is insecure regarding challenges to his or her sense of self. Interpersonal difficulties may be experienced during attempts to get others to submit to his or her opinions (Exner, 2003).

Color Projection (CP)

This unusual response indicates persons who deny unpleasant emotions by creating false or substitute emotions instead. Thus, they have difficulty dealing with negative feelings and modulating their emotions, and they bend or even distort reality as a means of adapting (Exner, 1993, 2003). This scoring category should be interpreted only in the context of other indicators for processing and expressing affect (see the Affect Section).

Ratios, Percentages, Derivations

The quantitative formulas used to develop the different ratios, percentages, and derivations provide a more in-depth and complicated portrayal of the relationships among the locations, determinants, contents, Populars, and special scores. These formulas provide some of the most important, reliable, and valid elements of interpretation. Their numbering and organization correspond with the numbers given to them in the previous listing of the quantitative formulas (see Structural Summary).

The Core Section

The Core section provides information on the person's dominant personality style, particularly focusing on the level of stress the person is experiencing and how effectively the style allows him or her to tolerate stress. Seven of the entries for the core section are frequency data providing summaries for total number of responses (R), animal movement (FM), inanimate movement (m), Achromatic color (C′), Shading—Texture (T), Shading—Dimension (V), and Shading—Diffuse (Y). Interpretive material for each of the last six categories can be found in previous sections; the first category, number of responses (R), is detailed in the following subsection. The nine quantitative formulas follow the interpretive material on number of responses.

Number of Responses Number of responses is not a quantitative formula (and is, therefore, not numbered). Rather, it is a simple sum of the total number of responses. In using Exner's set of instructions, the mean for the total number of responses for nonpatient adults is 22.32, with a standard deviation of 4.40. However, different

methods of administration can influence this number to a certain extent. For example, Ames et al. (1973) report an overall adult average of 26; S. J. Beck (1961) gives 32 for his adult mean; and both use instructions somewhat different from Exner's. Deviations from the normal range present the following possible interpretive hypotheses.

Low R (Adults, less than 17; Children, less than 15) A low number of R suggests defensiveness, constriction, organicity, depression, or attempted malingering (Exner, 1986). A clinician cannot confirm any of these hypotheses based solely on the occurrence of a low number of R, but they are raised as possibilities. To confirm these hypotheses, factors from both within and outside the test must be used. Protocols having fewer than 14 responses are not likely to be valid, and interpretations of the test, especially those based on the quantitative formulas, should be avoided (Exner, 1988, 1993). Examiners should either discard the protocol and interpret the verbal material clinically, or they should immediately retest the person and request that he or she include more responses.

High R (Greater than 33) A significantly higher-than-average number of responses suggests several possibilities, including an introversive character (Murstein, 1960; E. Wagner, 1971), above-average intelligence with a relatively high level of academic achievement (S. J. Beck, 1945; Goldfried et al., 1971), and a high degree of creativity (H. Adams, Cooper, & Carrera, 1963; Dana & Cocking, 1968). It can also suggest good ego functioning, including the ability to plan ahead, adequate impulse control, and the ability to tolerate stress (Goldfried et al., 1971; B. Klopfer & Davidson, 1962). Among persons with psychopathology, high R is found among manics and obsessive-compulsives (Alcock, 1963; S. J. Beck, 1945, 1951).

A high number of R is likely to alter the meaning of specific formulas or render them useless. A higher proportion of D and Dd responses can be expected because the number of W responses is usually exhausted sooner. Pure F responses also tend to increase in frequency, the number of Populars increases, and there are usually relatively more R for Cards VIII and X, thus elevating the Affective Ratio (Afr). Thus, interpretations based on the quantitative formulas derived from a high number of R should be treated with appropriate caution.

1. ***Lambda [L; (Pure F:Non-Pure F)].*** The Lambda index was developed by S. Beck (1961) as an improvement on the F% that had been used by other Rorschach systematizers. The earlier F% used the total number of R as the denominator, whereas the Lambda uses the total number of non-pure F. The Lambda ratio is used as an overall index of the degree of responsiveness versus lack of responsiveness to stimuli (Exner, 2003). Thus, persons can range from highly constricted and withdrawn to completely emotionally flooded by their responses to stimuli. The Lambda for nonpatients is between 0.11 and 2.33, with a mean of .60. In contrast to this are schizophrenics (.05 to 29.00), depressives (.08 to 15.00), and character-disordered persons (0.015 to 16.00; Exner, 1993, 2003). The important factor with these statistics is that psychiatric groups have a much wider range than normals. Thus, a maladjusted person may have a Lambda either greater than 0.99 or less than 0.32. The significance lies in Lambda's ability to provide specifics regarding the form this maladjustment takes. It is also

important to look at other information within the test, such as form quality and Experience Balance, to obtain a more complete conceptualization of the meaning of L.

High L (L, .99). Because pure F responses represent a withdrawal from experiencing a situation fully and an avoidance of perceiving all the possibilities that may be present, high L persons are likely to be conservative, insecure, and fearful of involvement (Exner, 1991, 1993). Such individuals have also been described as defensive, constricted (B. Klopfer et al., 1956; Z. Piotrowski, 1957), unimaginative (Alcock, 1963; Levi, 1976), and anxious (Riessman & Miller, 1958; J. Singer, 1960). Levitt and Truumaa (1972) have demonstrated the association of high L with depression, guilt, and an increased potential for suicide. Lambdas of 1.20 or greater are found in persons who have an excessive degree of affective detachment, often screening out relevant information (Exner, 1986, 1993). Thus, they avoid the complexities of a stimulus and often develop "tunnel vision" relating to certain ideas or perceptions. However, with adolescents, an interpretation that focuses on maladjustment should be made with caution because adolescents usually have a higher proportion of pure F responses (Ames et al., 1971; Exner, 1995).

Low L (L. < 32). A low Lambda generally indicates that the person becomes overinvolved with stimuli to the extent that affect disrupts cognitive functioning (Exner, 1993, 2003). Low L scorers have also been described as having inadequate control over their emotions; their frequent, impulsive acting-out results in difficulty maintaining satisfactory interpersonal relationships (Allison et al., 1968; Exner, 1993). Such people often have an impaired ability to attend to their environment (Alcock, 1963) and are often victims of their needs and conflicts (Exner, 1993). However, low Lambda might also be associated with persons who are achievement-oriented and who deal effectively with their environment (Exner, 1993). These characteristics are often suggested by other indicators in their protocols reflecting control and flexibility (average X + %, average number of Populars, good Organizational Activity, above-average W). This is consistent with the finding that increases in Lambda (along with decreases in es) have been associated with treatment improvement among children (Gerstle, Geary, Himelstein, & Reller-Geary, 1988).

2. ***Experience Balance or Erlebnistypus [EB; (M:C)].*** The Experience Balance formula or Erlebnistypus was originally devised by Rorschach and is the ratio between the sum of all M responses compared with the sum of all weighted color responses. Rorschach systematizers and researchers have come to view the Experience Balance ratio as the extent to which a person is internally oriented as opposed to being more externally directed and behaviorally responsive to outside stimuli. Although the EB ratio is usually relatively stable (Exner, Armbuster, & Viglione, 1978), it can temporarily change during times of stress or become more permanently altered during the course of successful psychotherapy (Exner, 1974; Exner & Sanglade, 1992). Although the EB ratio is usually stable for adults, there is considerable variability in children until mid-adolescence (Exner et al., 1985; Exner & Weiner, 1995). In an extensive literature review, J. Singer (1960) described the two sides of the ratio as representing dimensions of "constitutional temperament." These dimensions are introversives (higher M scores), who have a preference for internal experience, as opposed to extratensives (higher weighted C scores), who are more prone to activity and external expression. An introversive can more effectively delay his or her behavior, whereas the extratensive is

more emotional and is likely to discharge his or her affect into some form of external behavior. Both types respond differently to stress and to problem-solving tasks (Exner, 1978). It should be emphasized that, in their moderate forms, neither is any more or any less effective than the other, nor is either more prone to psychopathology (Molish, 1967; see also the interpretive meanings associated with M and C and the quantitative formulas dealing with either of these factors [EA, EBPer, D Score, Adjusted D Score, and W:M].)

Higher M (Introversives). Rorschach stated that persons with a relatively higher number of M responses were more oriented toward using their inner fantasy life. Thus, they are directed inward and use their inner experience to satisfy most of their basic needs. This is not so much an absolute necessity as it is a preference. In fact, these individuals may, on a more superficial level, even appear to be extraverted. Researchers have found them to be cautious, deliberate, submissive (Kurz, 1963; M. Rosenthal, 1962), and less physically active than persons scoring relatively higher on the C side of the ratio (Mukerji, 1969). They approach problem-solving tasks by internalizing the situation and mentally reviewing possible alternatives, and they engage in relatively few behaviors before reaching a solution (Exner, 1978).

Higher C (Extratensives). Persons who have relatively lower M responses and higher C responses (extratensives) tend to use external interactions as the most important means of satisfying their needs. They characteristically direct their energy toward the outside world. Extratensives are usually spontaneous and assertive, but also have difficulty delaying their responses (Alcock, 1963; Exner, Thomas, & Martin, 1980; J. Palmer, 1970). In children, higher C scores may represent a lack of self-assurance (Palmer, 1970). Extratensives are likely to approach problem-solving situations by experimenting with different behaviors (external trial and error) before achieving solutions (Exner, 1978).

M and C Equal (Ambitents). If M and C are equal, individuals are more likely to be flexible during interpersonal relationships, but they are also less sure of themselves during problem solving and tend to vacillate (Exner, 1978). They usually need to verify every sequence in the solution of a problem at hand, and they do not profit as much from mistakes as either the introversive or the extratensive does (Exner, 1978). Whereas the latter types are more sure of which response style to employ in approaching an ambiguous situation, ambitents have a liability when flexibility is required (Exner, 1978). Thus, ambitents tend to be less consistent and efficient than either introversives or extratensives. Unusually high scores on both M and C suggest a manic condition (S. J. Beck, 1960; J. Singer, 1960).

3. ***Experience Actual [EA; (M + C)].*** Whereas the Experience Balance ratio emphasizes the assessment of a person's type, the Experience Actual indicates the "volume of organized activity" (S. J. Beck, 1960). The M side of the formula shows the extent to which persons are able to organize their inner lives, and the C side indicates the extent to which emotions are available. The emphasis here is that both the M and the C represent deliberate, organized activity, which is contrasted with the disorganization associated with nonhuman movement (FM, m) and the responses related to the gray-black features of the blot (T, V, Y).

For the most part, the adult ratio between M and C is remarkably stable (Exner, 1993), yet the sum of M and C can sometimes fluctuate on a daily basis, which theoretically

parallels the effects of changes in mood (Erginel, 1972). After successful psychotherapy, particularly if long term, M and C typically both increase (Exner & Sanglade, 1992; Weiner & Exner, 1991), indicating a greater increase in the degree of organization of the person's inner life and an availability of more emotions. In fact, Exner (1974) found that EA increases significantly more for patients who improved in therapy than for those who showed little or no improvement. Furthermore, persons who underwent long-term insight-oriented treatment showed much more of an increase in EA than those in a treatment that emphasized a combination of support and environmental manipulation (Exner, 1974). This is consistent with the goal of insight therapy, which focuses on helping patients to understand and organize their internal resources. The mean changes for children show a gradual increase (rarely more than 0.5) with each year from the ages of 5 to 13 (Exner, 1993). Although brief retesting for children has shown good stability, long-term retesting (nine months or more) has resulted in wide fluctuations (Exner et al., 1985; Exner & Weiner, 1995).

4. ***Experience Pervasive (EBPer).*** Because Experience Balance (M:WSumC) is a somewhat crude indicator of how pervasive or dominant the introversive or extratensive style is, Experience Pervasive was designed as a more refined means of indicating how dominant one of the two styles is. Thus, it is an extension of the interpretations described in Experience Balance. It is calculated only when a clear style is indicated, and it takes on interpretive significance only when the value of one style over the other is 2.5 or greater. When this occurs, it clearly indicates that one of the styles is quite pervasive, perhaps to the point of suggesting rigidity in problem-solving style (Exner, 1993).

5. ***Experience Base [eb; (FM + m)/(Y + T + V + C′)].*** The Experience Base ratio was originally suggested by B. Klopfer et al. (1956) and later developed in its present form by Exner (1974, 1986). The nonhuman movement side of the ratio reflects tendencies to respond in ways that are not completely acceptable to the ego. These tendencies appear out of control, impinge on the individual, and are disorganized (B. Klopfer & Davidson, 1962). Although the tendencies and feelings may have originally been produced by outside sources, the resulting internal activity is not in the person's control. The opposite side of the ratio, which is a sum of the responses relating to the gray-black features of the blot, is a reflection of the pain and disharmony the person is feeling as a result of unresolved stress. The eb ratio indicates which of these two areas of functioning is more predominant. If the eb is small on both sides, it suggests that the person is not experiencing very much pain and that his or her needs are well organized. Usually, the values on either side of the ratio range between one and three for nonpatients. If either side becomes greater than five, its interpretive meaning becomes more clear. (See also the additional interpretive meanings associated with material from the left side of the ratio [FM and m] and the right side [Y, T, V, and C].)

6. ***Experienced Stimulation [es; (FM + m) + (C′ + T + Y + V)].*** Experienced Stimulation is the sum of the nonhuman movement responses and all responses relating to the gray-black features of the inkblot. These are all responses reflecting that the person's functioning is disorganized and that forces are acting on the person and he or she feels they are beyond control. Thus, the es sum is an index of a person's degree of disorganization and helplessness. Persons scoring high on es have a low frustration tolerance, and it is difficult for them to be persistent, even in meaningful tasks (Exner, 1978).

Important information can be obtained by comparing the amount of organization the person has (as represented by EA) with how much chaos and helplessness he or she experiences (as represented by es). Normal populations usually have a higher EA than es, whereas psychiatric populations have a higher es than EA (Exner, 1974). Exner (1978) has suggested that the ratio between EA and es can provide an index of the degree to which a person can tolerate frustration. Difficulty in dealing with frustration would primarily result from high-scoring es persons' having a limited ability to process and mediate cognitive information (Wiener-Levy & Exner, 1981). As would be expected, a correlate of successful psychotherapy is that there is a decrease in es and a corresponding increase in EA, which suggests that at least some of the patient's activity has become more organized (Exner & Sanglade, 1992; Gerstle et al., 1988; Weiner & Exner, 1991). This was supported by Exner (1974), who found that subjects rated as unimproved after therapy also showed little change in that their es still remained high in relationship to EA. In another study, Exner (1974) demonstrated that most persons in successful insight therapy had an increase in EA compared with es. This suggests that patients in successful insight therapy were able to either neutralize or reorganize the forces that were "acting on" them. In contrast, therapy emphasizing support or environmental manipulation produced no or little change in the es:EA ratio.

High es (11 or higher). High scores suggest low frustration tolerance, difficulty following through on tasks, disorganization, distractibility, and a sense of helplessness (Exner, 1978, 1986).

7. ***D Score (D; EA − es).*** The D Score is a further measure of the client's ability to tolerate stress. It is essentially a means of evaluating the degree of available resources the person has (EA) versus the amount of disorganized events that are occurring beyond the person's control (es). For example, veterans diagnosed with PTSD have been found to have low D Scores (Weiner, 1996).

Low D Score (−1 or lower). A low D Score indicates that es characteristics have relatively more weight than EA. The person is likely to feel overwhelmed and unable to deal with complex or ambiguous situations. His or her thoughts, affects, and behaviors might be impulsive and poorly focused (Exner, 1993, 1995). As the D Score becomes progressively lower, this trend is likely to become increasingly stronger. A sense of being overloaded, easily distracted, and limited psychological resources to deal with stress are characteristic.

High D Score (+0 or higher). A D Score of 0 or higher indicates that the client can adequately deal with the current level of stress. Even if experiencing stress (high es), the relatively higher D Score indicates that he or she has adequate resources to cope effectively with this stress (Exner, 1993, 1995).

8. ***Adjusted es (Adj es).*** Because es includes measures of current stimuli impinging on the person (m and Y), a different, adjusted es that excluded m and Y was developed. This adjusted es represents the more chronic (rather than fluctuating) condition of the person (Exner, 1993). Thus, persons scoring high are likely to feel chronically overstimulated (i.e., racing thoughts, insomnia) and have difficulties organizing their thoughts. However, the main purpose of calculating Adjusted es is to enable the calculation of the Adjusted D Score.

9. ***Adjusted D Score (Adj D).*** Because the D Score includes measures of the current capacity to deal with stress, it may not provide a measure of the person's usual ability

to modulate and control his or her behavior. This issue is particularly likely to be present for clients referred for evaluation, because the events surrounding a referral usually involve psychosocial difficulties. These situational uncontrollable stressful events are expressed on the Rorschach (and in the D Score) by the presence of m and Y responses (McCown et al., 1992). Adj es has had m and Y subtracted from it, so it theoretically removes the influence of current environmental stressors. What remains in the Adjusted D Score is a measure of the person's typical or usual capacity to tolerate stress and to control behaviors (Exner, 1993, 1995).

Low Adj D (−1 or lower). These persons have fewer than average resources to adequately cope with stressful situations. They function best in routine and predictable situations. Adapting to new situations presents difficulties in that they are prone to become distracted, disorganized, and impulsive. These trends are strengthened with progressively decreasing scores on Adj D.

High Adj D (+1 or higher). High scores on Adj D indicate a good ability to deal with stressful situations but not necessarily a well-adjusted person. For example, antisocial or paranoid personalities might have intricate systems of dealing with stress that are quite effective at reducing their anxiety levels; yet, they are clearly not well-adjusted persons. In addition, treatment might present difficulties. They may use their somewhat limited resources to distance themselves from the types of experiences that might potentially stimulate increased growth and awareness (Exner, 1993). In contrast, a certain amount of distress can be useful in motivating the client to change.

The Ideation Section

The Ideation section focuses on information related to how the client imposes meaningful organization onto his or her perceptions. It includes three quantitative formulas (two ratios and an index) and frequency data for M–, M, number of Level 2 responses, WSum6, and M with no FQ (see the interpretation for each of these frequencies under the listings for human movement and Special Scores).

1. ***Active*:*Passive Ratio (a*:*p).*** Individuals who have a distinctly higher number of passive responses are likely to be correspondingly more passive in other situations. In contrast, a clearly higher number of active responses indicates a person who is more active in terms of thoughts and behaviors (see also the interpretation for M^a:M^p; Exner, 1974, 1993). However, the contrast or magnitude of differences must be quite clear, as indicated by one of the following conditions: (a) "sum of the values in the ratio is four and one value is zero"; (b) "values in the ratio exceed four, and the value on one side of the ratio is no more than twice that of the other"; or (c) "ratio exceeds four, and the value on one side is two to three times greater than the value on the other side" (Exner, 1993, p. 475).

2. ***M Active*:*Passive Ratio (M^a*:*M^p).*** A further refinement of the a:p is to consider only the proportion of active and passive responses for human movement scorings. If the summed value of passive Ms (M^a) is greater than active Ms (M^p), it suggests a generally more passive orientation. For example, therapists' ratings of clients with a greater number of passive Ms indicated that they made more requests for directions, seemed more helpless, and exhibited a relatively high number of silences (Exner,

1978). In addition, their daydreams had more passive themes (Exner, 1974) as did their TAT story endings (Exner, 1993).

3. ***Intellectualization Index [2AB + (Art + Ay)].*** Earlier research indicated that the presence of three or more summed scores for Abstraction (Ab) and Art (Art) suggests an excessive use of intellectualization (Exner, 2003). Both obsessives and paranoid schizophrenics were often found to have more than three combined Ab and Art frequencies in their protocols (Exner, 1986; Exner & Hillman, 1984), and both these groups are likely to use an intellectual approach to distance themselves from their emotions. This is in contrast to other patients and nonpatients who typically reported an average of approximately one per protocol (Exner, 1986).

High Intellectualization Index (Six or More). Although scores of four or more can indicate a trend toward intellectualization, scores of six or more are usually necessary to more strongly suggest this mode of defense. High scorers are likely to neutralize their emotions through analyzing things from an intellectual perspective. This might serve to deny or conceal the impact of affect. Their method of dealing with emotions is typically quite circumspect and possibly unrealistic. Although this might provide them with a certain degree of control for moderate levels of affect, much higher levels of affect are likely to overwhelm them, quite possibly resulting in disorganization (Exner, 1993).

Affect Section

The Affect section provides information on how the person modulates and expresses affect. Because affect is most directly expressed on the Rorschach through color, the different frequencies and formulas are concerned with the various combinations of color with other types of Rorschach responses. Specifically, this section includes frequencies for Pure C, white spaces (S), color projection (CP), and three quantitative formulas.

The sum of C and Cn responses provides an index of the degree to which a person is likely to be overwhelmed by affective impulses. Among nonpatient adults, it is rare to have any C or Cn responses occurring in a protocol (M = 0.12, SD = 0.43), but this increases slightly for patient groups (see discussion in the section on interpretation of color). The degree to which a person uses white spaces (S) has been associated with the person's negativism, means of handling anger, and amount of oppositional tendencies (see Interpretation section on white spaces). Color projection (CP), a rare response included as a Special Score, relates to a tendency for the individual to substitute alternative emotions in place of unacceptable unpleasant ones (in the Interpretation section, see the discussion of color projection [CP]).

1. ***Form-Color Ratio [FC/(CF + C)].*** The ratio of form-dominated color responses to color-dominant responses provides a measure of the degree of control a person has over his or her impulses (also check D Score for a tendency to become overwhelmed by stress). If form is predominant (1.5 to 2.5 times greater), it suggests the person has good control over his or her impulses and experiences satisfying interpersonal relationships (Exner, 1969, 1974; B. Klopfer & Davidson, 1962). Exner (1978), for example, has found that schizophrenics who have FC responses greater than CF + C have a better response to psychotherapy and less likelihood of relapse. The high form suggests they can integrate an accurate, reality-oriented interpretation into their perceptions.

However, if no or very few color-dominant responses (no CF + C) are present, the person will be overly constricted and have little contact with his or her emotions (Exner, 1978, 1993). This is consistent with the finding that most psychosomatic patients, who are typically constricted, had ratios of 4:1 or greater (Exner, 1993). If the CF + C side of the ratio is relatively high (1:1), it suggests a weak control over a person's impulses, which may be accompanied by impulsiveness (Pantle et al., 1994) or aggressive acting out, perhaps consistent with a narcissistic personality (Exner, 1969; B. Klopfer & Davidson, 1962). The perception of both internal and external events is typically distorted and inaccurate, as are the responses to these events (Exner, 1974). The number of pure C responses increases with pathological groups, as indicated by only 7% of nonpatients giving pure C responses in contrast to 45% of depressives, 32% of schizophrenics, and 27% of character-disordered patients (Exner, 1993).

2. ***Affective Ratio [Afr; (R for Cards VIII, IX, X)/(R for Cards I-VII)].*** Because the last three cards are chromatic and the first seven are primarily achromatic, the Affective Ratio indicates the extent to which affect (color) makes an impact on the person. Nonpatient adults usually show a mean Afr of .67 (SD = 0.16). However, it is relevant to consider Afr in the context of EB. Introversives (higher M side of EB), who primarily direct their experience inward, have Afr ranges between .50 and .80. In contrast, Extratensives (higher C side of EB) have Afr ranging between .60 and .95 (Exner, 1993, 2003). This means that it is useful to take EB scores into account when judging whether an Afr is high or low. Although the mean Afr for patient groups was not very different from that for nonpatients, the range was much higher for patients and the distribution was bimodal. This higher range among patient groups is consistent with the view that they are more likely to have difficulties with either undercontrolling or overcontrolling affect (Exner, 1993, 2003).

High Afr (Greater Than .85). A high Afr indicates an overresponsiveness to affect (Exner, 2003), reflecting that the person is more receptive to emotional inputs and more likely to respond immediately rather than delay behavior (Exner, 1995, 2003). It is also important to evaluate the FC/(CF + C) proportion to assess the degree of control the person has over his or her emotions. In other words, the Afr measures responsiveness, interest in and degree of affect, whereas the FC/(CF + C) indicates the ability to control what affect is present.

Low Afr (Less Than .53). Persons with low Afr scores tend to withdraw from their emotions and, if they have an unusually low Afr, may attempt to exert an extreme amount of control over their affective responses (Exner, 1993, 1995, 2003). (To assess the possibility that they avoid emotions through intellectualization, note their scores on the Intellectualization Index.)

3. ***Complexity Index (Blends:R).*** Approximately 20% of all Rorschach responses involve blends. To create a blend response, the person must appreciate the complexity of the inkblot, which requires both analysis and synthesis. Exner (2003) has pointed out that the pure F response is the exact opposite of a blended response in that pure F requires attention to only the most simple, straightforward aspect of the stimulus. Usually, there are one or more blends in a person's protocol. A complete absence of blends suggests narrowness and constriction. This is consistent with the finding that blends are less frequent in the protocols of depressives and persons with below-average intelligence (Exner, 1993, 2003). In contrast, an extremely high number of blends (eight or

more) suggests an unusual amount of complexity, to the extent that the person may be overly burdened (Exner, 1993, 2003).

A thorough interpretation of blends also requires an understanding of their qualitative aspects. For example, a blend that includes color-dominated determinants implies that the person might be easily overwhelmed by affect, whereas the opposite would be true if the blend were form-dominated. The color-shading blend (combining color with C′, Y, T, F, V) implies concern with painful, irritating, confusing emotional experiences, and it is associated with the protocols of depressives. Exner and Wylie (1977) found a moderate correlation with attempted suicide. Accordingly, this blend was included as one of several variables in Exner's (1993) Suicide Constellation. However, the presence of color-shading blends does not seem to be a sufficiently accurate predictor of suicide when used as a single sign (Hansell et al., 1988).

The Mediation Section

The Mediation section uses a series of indicators to measure the extent to which the client is oriented toward making conventional, acceptable responses versus more unique ones. If either one of these directions is extreme and rigid, it suggests difficulties in adapting. This section includes simple frequencies for the total number of Populars and negative white space responses (S–; see previous interpretation for Populars and white space responses) along with the five percentages described next.

1. ***Conventional Form (X + %).*** X + % includes the form quality of all the responses in a protocol and, as such, tends to be less subject to distortions than F + % (see next item). The X + % is essentially an indicator of the degree to which a person perceives things in a conventional, realistic manner. Most normal adults have an X + % of 77% (SD = 9%; Exner, 2003). Normal children have a slightly lower mean, ranging between .67 and .78 (Exner & Weiner, 1995). An extremely high percentage (greater than 90%) means that persons perceive their world in an overly conventional manner, to the extent that they might sacrifice their individuality. They are likely to be hypernormal, inflexible, rigid, and overly conventional (Exner, 1993, 2003). This is further supported by, and is consistent with, an elevated number of Populars. In contrast, lowerings in X + % (less than 70%) suggest persons who perceive their world in an unusual manner. This might be simply because they are highly committed to their individuality or, particularly if X + % is unusually low, it might suggest serious psychopathology. For example, schizophrenics have a mean X + % of only 40% (Exner, 1993).

2. ***Conventional Pure Form (F + %).*** F + % assesses the same dimension as X + % but is limited to a narrower number of responses because it involves only pure F responses rather than other scoring categories (C′, Y, T, and V) that might have been combined with F. Thus, interpretation is similar to the interpretation of X + % but should be done more cautiously. It reflects a person's respect for the conventional aspects of reality and perceptual clarity. The Exner (1993) norms indicate schizophrenics have an F + % of only 42%, in contrast to the average of 71% among normals. In general, a low F + % might suggest limited intellectual endowment (S. J. Beck, 1961), organic impairment (Reitan, 1955b), or schizophrenia (S. J. Beck, 1968; T. Kahn & Giffen, 1960).

3. ***Distorted Form (X − %).*** In contrast to X + % (and F + %), X − % is a direct index of the degree to which a person has distorted perceptions of reality. The higher

the X – %, the more likely that the person has a significant level of impairment. For example, moderately high percentages (X – % = 20%) are found for depressives, and percentages of 37% are characteristic of schizophrenics (Exner, 1993). Any percentage above 20% suggests that the person will have difficulty, because he or she has poor ties with reality and difficulty developing accurate abstractions.

4. ***White Space Distortion (S – %).*** Sometimes, X + % and F + % can be low and it might then be assumed that this is a result of a high number of form minus responses. This assumption might then result in incorrect interpretations. One way of checking for this difficulty is to note the percentage of minus responses for the white space (S–). Instead of suggesting the sort of distortions suggestive of schizophrenia (see interpretations for F + % and X + %), a low S – % might be caused by strong negativism or anger (Exner, 1993, 2003).

5. ***Unusual Form (Xu%).*** Xu% also provides a check for potentially incorrect interpretations derived from low X + % or F + % scores. There might be cases in which X + % and F + % are low primarily as a result of a large proportion of unusual form (Fu) responses. Fu responses are unusual, but they still do not violate reality in the way that minus responses do, and thus they do not reflect severe pathology. In fact, a few Fu responses in a protocol can be a healthy sign that the person is capable of seeing his or her world in a novel manner. However, an overabundance of Fu responses suggests the person is highly committed to an unconventional orientation (Exner, 1993, 2003). Unless the environment is highly tolerant of such an orientation, he or she is likely to have numerous conflicts and confrontations.

The Processing Section

In addition to understanding clients' ideation and mediation, it is also important to assess the quality and efficiency by which they process information. Relevant frequency data are the overall amount of Organizational Activity (Zf; see interpretation under Organization Activity), Perseveration (PER), Developmental Quality+ (DQ+), Developmental Quality v (DQv), and the following three ratios:

1. ***Economy Index (W: D: Dd).*** The W: D: Dd ratio compares the degree to which an individual attempts to create a more challenging response that requires a high degree of organization and motivation (W), rather than choosing a less demanding and easily perceived area (D or Dd). Normals and outpatients usually have a W: D ratio of 1: 1.2 or even 1: 1.8 (Exner, 1993). If a person includes a relatively large number of D responses, it suggests that he or she takes the least challenging and possibly least productive way out of a conflict situation. It could be assumed that his or her characteristic way of dealing with ambiguity is to withdraw from it and focus on the obvious. If W is predominant, the person is perhaps overdriven in his or her attempts to organize perceptions. If, with a high W, both the W and the D responses are of poor quality, it suggests that a person is withdrawn and unrealistically striving for perfection (Exner, 1974). However, when W and D responses are both of good quality, they more likely represent the successful intellectual efforts of a creative person (Exner, 1974).

2. ***Aspirational Ratio (W: M).*** The W: M ratio is a rough formula that, at the present time, is somewhat lacking in research. It can be generally understood by reconsidering that the W response is an indicator of the degree to which subjects aspire to

effectively organize and conceptualize their environments. It is an effort to encompass and include a number of different details in one coherent response. However, determining whether subjects have the resources to actually accomplish an effective organization depends also on M. Although M represents the degree of investment subjects have in their fantasy lives, it also suggests how effectively they can bridge their inner resources with external reality and perform abstract thinking. Thus, the W: M ratio provides a rough comparison between a person's aspiration level, as represented by W, and his or her actual capability, as represented by M (Exner, 1993, 2003). Because introversives have higher M values than extratensives and ambitents, the relative value of EB needs to be taken into account in designating high or low W: M ratios. A high aspirational level is indicated if the W side of the ratio is greater than the following values: introversives, 1.5: 1; ambitents, 2.2: 1; extratensives, 3: 1 (Exner, 1993). However, scores with extremely high W components are common in children, which is consistent with the observation that children often underestimate the actual effort required to accomplish a goal (Exner, 1993, 1995). On the other hand, ratios where the right side (M) is clearly lower than the left (0.5 : 1 for introversives, and 1 : 1 for extratensives and ambitents) suggest that these persons are extremely cautious and conservative in defining achievable goals (Exner, 1993, 2003). Their motivation to achieve might be low, which would involve their being cautious (not wishing to fail), conservative in defining their objectives, and economical in their expenditure of energy.

3. ***Processing Efficiency (Zd).*** Although the frequency of Organizational Activity (Zf) along with the Economy Index (W : D : Dd) and the Aspirational Ratio (W : M) provide information on the motivation and effort that persons place into their perceptions, these indicators do not provide information related to quality or accuracy. In contrast, the Processing Efficiency (Zd) score provides an index not only of effort but also of ease and accuracy of processing. Individuals scoring high on Zd are considered to have an overincorporative style; they invest more effort and are more accurate in their perceptions and conclusions. This seems to be an enduring traitlike feature. In contrast, low scorers have an underincorporative style, which means that they process information in a more haphazard style, often neglecting relevant bits of information. This characteristic seems more amenable to change, as indicated by moves to a more overincorporative style following psychotherapy (Exner, 1978). A review of research on Zd (Exner, 1993) indicates that, consistent with theory, overincorporators (high Zd) have more extensive eye-scanning, make fewer errors on games, and are less likely to make guesses related to requests for factual information. In contrast, underincorporators (low Zd) make fewer eye movements while scanning, are more likely to make errors on games, and are more likely to make guesses related to factual information. Among children, low Zd scores occur among those diagnosed as hyperactive.

High Zd (> + 3). High scorers on Zd are more likely to be obsessive or perfectionistic but can also efficiently and accurately process information. They are likely to take care with their perceptions and continually check for accuracy. They exert more effort in information processing and are more confident in their abilities.

Low Zd (< – 3). Persons scoring low on Zd are more likely to be haphazard and to make impulsive decisions without fully taking into account all relevant aspects of a situation. Compared to high Zd scorers, they do not invest as much effort into actively

working with their perceptions. They typically are uneasy with their information-processing ability and may question their efficiency at perceiving, integrating, and responding to information.

The Interpersonal Section

Although the Rorschach does not obtain information regarding a person's actual environment or the other persons in that environment, it does provide information related to needs, attitudes, behavioral response sets, and coping styles, all of which are relevant to interpersonal relationships. The Interpersonal section lists several measures relevant to these domains. The person's degree of cooperation with others can be noted through their number of Cooperative Movements (COP). Similarly, Aggressive Movements (AG) provides an index of interpersonal aggression, and a high amount of Food Contents suggests dependency. Additional useful indicators of interpersonal relations include sum of pure H, number of Perseverations (PER), ratio of Good to Poor Human Representations (GHR:PHR), sum T, and active:passive responses (see interpretations under each one of these categories). The following two formulas can also be useful in assessing the extent to which a person is interested in people as opposed to being isolated.

1. ***Interpersonal Interest [H + (Hd) + Hd + (Hd)].*** The Interpersonal Interest ratio index merely adds up the total amount of human content. As such, it represents the degree to which a person is interested in people.
2. ***Isolation Index (Isolate:R).*** Exner (1986) points out that the five contents (Botany, Clouds, Geography, Landscape, and Nature) used to develop the Isolation Index are all "nonhuman, nonsocial, inanimate, and usually static objects" (p. 406). If a high proportion of these contents (index score of .25 or greater) occurs in a person's protocol, it suggests the person may be withdrawn or alienated, or may at least have some difficulties related to social isolation (Exner, 1993). This seems to be true for children, adolescents, and adults (Exner, 1986, 1995). However, these interpretations should not necessarily take on a pathological bias. A high score might merely represent less interest in people rather than a negative rejection and alienation from them.

Self-Perception Section

The Self-Perception section includes information relevant to the relative assets and limitations of the clients as seen by the clients themselves. The following entries are simply frequency tallies: Fr + rF, Form Dimension, sum of Morbid content, and Anatomy/X ray responses, and sum V (see interpretations under relevant sections). The ratio of Pure H:(H) + Hd + (Hd) compares the amount of Pure Human responses with mythical/fictional and part human responses. Two of the human categories on the right side of the ratio relate to fictional/mythical descriptions. As such, they can be considered to represent the extent to which the individual bases his or her perceptions on real versus imaginary aspects of people. Adult and adolescent nonpatients usually give more Pure Human responses than (H) + Hd + (Hd) at a rate of approximately 3:2 (Exner, 1993, 1995, 2003). However, the means for the ratio are different for introversives (3:1) than for either extratensives or ambitents (1.3:1). In contrast, schizophrenics see a much higher proportion of fictional/mythical and part human responses (1.5:2; Exner, 1993). This low a level of Pure Human responses suggests that they are

working from an unrealistic perception of themselves and others. The Self-Perception section also includes one quantitative formula:

1. ***Egocentricity Index [3r + (2)/R].*** The Egocentricity Index (EI) provides information related to whether the client has a sense of self-worth and further relates this to the extent that he or she is absorbed with self.

High EI (>.49). A certain level (index level of .40 to .45) of self-focusing and self-concern is associated with positive self-esteem (Exner, 1993). However, index scores above .49 suggest that the person has an overinflated sense of self-worth, which reflects underlying dissatisfaction. This may be expressed in part by neglect of aspects of the external world.

Low EI (<.32). In contrast to high EI, scores below .32 indicate that the person's self-worth is quite negative and he or she feels conflicted regarding a self-image. This may lead to mood fluctuations along with dysfunctional behaviors (Exner, 1993).

Special Indices

In an attempt to increase the robustness and validity of various combinations of Rorschach measures, six Special Indices have been developed based on a composite of scores. For example, a number of different indicators of schizophrenia are found throughout the Rorschach. These include a high number of X – % or M – %, the presence of one or more Level 2 Fabulized Combinations (FAB2), and a high WSUM6. These, along with several other indicators, were combined to form the Perceptual-Thinking Index (PTI). Some research has found that this can discriminate schizophrenia better than any of the single scores (see Exner, 1991, 1993, 2003). A similar strategy was used for the other Special Indices. Collectively, they help to form a nucleus of indices to help with more specific types of diagnostic conditions. Exner (1993, 2003) and the commercially available scoring forms have included Constellation Worksheets for calculating whether the Special Indices are positive (also see Exner, 2003; Table 10.5). Caution should be used in interpreting these indices because most have either equivocal or minimal research to support them (see Archer & Krishnamurthy, 2003).

1. ***Perceptual Thinking Index (PTI).*** The PTI is a revision of the earlier Schizophrenia Index (SCZI). PTI has the advantage of more accurately identifying persons with thought disorders (Exner, 2003; S. Smith et al., 2002). As the name suggests, it is not designed to diagnosis schizophrenia but more to assess the array of disorganized or unusual thought processes that may occur with schizophrenia or other forms of thought disorders. It should also be considered as rating a person on a continuum of thought disturbances rather than being designed to place a person in a certain category (diagnosis).

2. ***Depression Index (DEPI).*** A DEPI value of 4 raises the possibility that the client is experiencing some depressive symptoms—fluctuations in moods, a sense of dissatisfaction, pessimism, and some mild vegetative symptoms (fatigue, insomnia, slowed thinking, anhedonia). Scores of 5, 6, or especially 7 are far more definitive and strengthen the likelihood of an affective disorder as reflected by an intensification of these symptoms (Exner, 1991, 1993). However, the diagnosis of a specifically depressive disorder may not be warranted because depression is generic to a wide variety of disorders, particularly many of the personality disorders and schizophrenia. In addition, the term *depression*

might be used to describe people who are emotionally distraught or are pessimistic, self-defeating, and lethargic, as well as those who feel a sense of futility when attempting to function competently in a complex society (Exner, 1993). Depressive symptoms as measured by DEPI may, therefore, relate to both a wide number of types of people and a wide range of possible diagnoses. This index should be used with considerable caution because it has been identified by a number of authors as having questionable validity, especially with adolescents (Archer & Gordon, 1988; Ball et al., 1991; Jorgenson et al., 2000; G. Meyer & Archer, 2001).

3. ***Coping Deficit Index (CDI).*** Clients with scores above 4 or 5 on the CDI are likely to have unsatisfying and somewhat meaningless interpersonal relationships, largely because they find it difficult to effectively deal with everyday requirements (Exner, 1993). Their histories typically include social ineptness, poor success in interpersonal relationships, and times when they have felt overwhelmed by interpersonal demands. Effective moderate- to long-term psychotherapy was found to result in decreases in CDI (Exner & Sanglade, 1992; Weiner & Exner, 1991).

4. ***Suicide Constellation (S-CON).*** The Suicide Constellation comprises 11 variables that collectively are intended to detect persons at risk of attempting suicide. Retrodictive studies indicate that, using a cutoff score of 8, 80% of suicide attempters were accurately identified (Exner, 1986, 1993). However, caution should be exercised in making final decisions. Some clients were incorrectly identified as not being suicidal and yet they later made attempts (false negative rate = 15%). Among depressed populations, a number of clients were incorrectly identified as being at risk of suicide when there was actually no or little risk (false positive rate among depressives = 10%; Exner, 1993).

5. ***Hypervigilance Index (HVI).*** Originally, a series of indicators was isolated from patient protocols that seemed to differentiate paranoid-type patients (paranoid schizophrenics) from other patient groups. This was partially successful in that paranoid schizophrenics and paranoid personalities were correctly identified (88% and 90%, respectively; Exner, 1993). On further investigation, it was found that HVI related more to the hypervigilant aspect of the paranoid style rather than paranoia itself. Thus, persons with positive indicators on HVI are likely to place a large amount of effort into maintaining a high state of preparedness. Motivating this is a sense that they mistrust their environment and experience a chronic sense of vulnerability (Exner, 1993). Before initiating behaviors, they carefully think through why and how they should express them. They are likely to be quite guarded regarding closeness in relationships and initially respond to efforts at closeness with apprehension. As a result, they allow themselves to be close with others only if they feel in control. They are generally quite concerned with issues not only of emotional closeness, but also of personal space in general (Exner, 1993).

6. ***Obsessive Style Index (OBS).*** The Obsessive Style Index was developed by examining the records of clients who had been formally diagnosed as obsessive-compulsive to determine which Rorschach characteristics could distinguish them from other groups. Five characteristics were isolated and, using the designated criteria, they correctly identified 69% of obsessive-compulsives (Exner, 1993). If OBS is positive, it suggests persons who are perfectionistic, indecisive, and preoccupied with details, and who experience difficulty expressing emotion. They are likely to be cautious,

conservative, conforming, and conventional (check for high Populars). They process information extremely methodically and, when using the Zd Index definition, are likely to be overincorporators (see interpretation for Zd). However, a positive index does not necessarily indicate psychopathology; rather, it shows a style of approaching the world and processing information. If this style is overly rigid, it can become dysfunctional, particularly when the person is under pressure or is required to achieve goals within a limited time (Exner, 1993).

RECOMMENDED READING

Exner, J. E. (2000). *A primer for Rorschach interpretation.* Asheville, NC: Rorschach Workshops.

Exner, J. E. (2001). *A Rorschach workbook for the comprehensive system* (5th ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E. (2003). *The Rorschach: A comprehensive system. Volume 1: Basic foundations.* (4th ed.). New York: Wiley.

Exner, J. E., & Weiner, I. (1995). *The Rorschach: A comprehensive system. Volume 3: Assessment of children and adolescents* (2nd ed.). New York: Wiley.

Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment, 13,* 486–502.

Chapter 11

THEMATIC APPERCEPTION TEST

The Thematic Apperception Test (TAT) is a projective technique consisting of a series of pictures. The examinee is requested to create a story about what he or she believes is occurring in the situations or events depicted by the pictures. The test was originally published by Murray and his colleagues at the Harvard Psychological Clinic in 1938. Murray (1943) describes the TAT as a "method of revealing to the trained interpreter some of the dominant drives, emotions, sentiments, complexes, and conflicts of personality. Special value resides in its power to expose underlying inhibited tendencies which the subject is not willing to admit, or cannot admit because he is unconscious of them" (p. 1). It differs from projective drawings or inkblot-type tests such as the Rorschach or Holtzman in that the TAT cards present more structured stimuli and require more organized and complex verbal responses. In addition, the TAT relies on more qualitative methods of interpretation and assesses the "here and now" features of an individual's life situation rather than the basic underlying structure of personality. It has often served as a model for the development of similar techniques. Past and current surveys of test use patterns indicate it is in the top seven most frequently used tests by professional psychologists (Camara et al., 2000).

The TAT materials consist of 20 cards with ambiguous pictures on them. The examinee is instructed to make up a story that includes what is occurring in the picture: the thoughts and feelings of the characters, the events that led up to the situation, and the outcome of the story. The examiner can interpret the responses either quantitatively (using rating scales to measure intensity, duration, and frequency of needs) or qualitatively (evaluating the story themes using clinical judgment). The results are typically used to supplement other psychological tests because the TAT produces not only highly rich, varied, and complex types of information, but also personal data that theoretically bypass a subject's conscious resistances.

HISTORY AND DEVELOPMENT

The TAT was first conceptualized in a 1935 article by Christina Morgan and Henry Murray but was more fully elaborated in 1938 and 1943 (Gieser & Stein, 1999). Administrators were instructed to give all 20 cards in a given sequence in two separate sessions, which, in total, could last up to two hours. The basic assumption was that unconscious fantasies could be revealed by interpreting the stories subjects told regarding ambiguous pictures. Examiners potentially gained access to things that a client was either unwilling to tell or unconscious of. Initially, it was believed that the material derived from the test

could serve as an "X ray" of personality and would reveal basic themes that might otherwise take months of psychoanalysis to understand. The test immediately had an enthusiastic reception and quickly became used as both a clinical instrument and a research tool. By 1950, several books and more than 100 articles were published either about or using the TAT. The early research studies using the TAT investigated areas such as social attitudes, delinquency, abnormal personality, and variations in the use of language. By the late 1940s, many clinicians were using a limited number of cards and abbreviated scoring systems to reduce the time for administration and scoring. These different TAT systems were elaborated in Shneidman's (1951) *Thematic Test Analysis.* By 1971, more than 1,800 articles had been written based on the TAT.

Despite continuing extensive research, the test is still not considered to have achieved a degree of standardization comparable to the MMPI/MMPI-2 or WAIS-III. There is no clear, agreed-on scoring and interpretive system, and controversy regarding the adequacy of its reliability and validity is ongoing. Most clinicians vary the methods of administration, especially regarding the number, sequence, and types of cards that are given (Gieser & Stein, 1999). As a result, the TAT is considered a highly impressionistic tool, with interpretation frequently coming from a combination of intuition and clinical experience. Yet, the TAT continues to be extremely popular and currently ranks as the sixth most frequently used test by clinical psychologists (Camara et al., 2000). Fully 63% of psychologists reported using it with adolescent clients (Archer et al., 1991). Clinical psychologists have made it the second most frequently recommended projective test for clinical psychology trainees' competence (Watkins et al., 1995). Furthermore, it has been used in all the European countries, India, South Africa, China, South America, Asia, and the Soviet Union (see Bellak & Abrams, 1997). The TAT (or TAT-type tests) has also been found to be the most frequently used assessment device for cross-cultural research (Dana, 1999; Retief, 1987).

A number of researchers were dissatisfied with the TAT because they wanted to study different populations (children, the elderly, minorities) and specific problem areas (frustration, stress, social judgment), or because they felt that the TAT produced negative, low-energy stories. These concerns stimulated numerous variations. The most common is the Children's Apperception Test (CAT; Bellak, 1954, 1986, 1993; Bellak & Abrams, 1997) designed for children between the ages of 3 and 10. Only 10 cards are given, and animals are depicted instead of humans. The rationale was that because children have shorter attention spans, they need fewer cards. It was also believed that they could more easily identify with pictures of animals than with pictures of humans. Subsequently, another version of the CAT was developed depicting humans instead of animals (CAT-Human or CAT-H). The Gerontological Apperception Test (Wolk & Wolk, 1971) and the more frequently used Senior Apperception Test (SAT; Bellak, 1975, 1986, 1993; Bellak & Abrams; 1997, Bellak & Bellak, 1973) are designed for elderly populations and show pictures of elderly people involved in scenes more likely to concern them, such as depictions of loneliness and family conflicts. The Tell Me A Story Test (TEMAS; Costantino & Malgady, 1999; Costantino, Malgady, & Rogler, 1988), designed for use with child and adolescent minorities, includes 23 cards depicting Hispanic, African American, and Asian American characters in situations of interpersonal conflict (Costantino & Malgady, 1999). There is also a parallel version for nonminorities. Scoring is made for nine different personality functions (aggression, anxiety, and so on), and the scores have been

found to effectively discriminate between minority outpatients and minority normal school children (Costantino, Malgady, Casullo, & Castillo, 1991; Costantino, Malgady, Rogler, & Tsui, 1988; R. Flanagan & Guiseppe, 1999) as well as nonminority normals and clinical groups (Costantino, Malgady, Colon-Malgady, & Bailey, 1992). The TEMAS is probably the best constructed and most psychometrically sound TAT variation to date (see R. Flanagan & Giuseppe, 1999).

Several TAT-type tests have been designed to study specific problem areas. The Rosenzweig Picture Frustration Study (Rosenzweig, 1976, 1977, 1978) was designed to more fully understand how persons perceive and deal with frustration. The Stress Tolerance Test is an older test that may begin to be used more frequently again in understanding how a subject responds to stressful scenes of combat (Harrower, 1986). More recently, K. Caruso (1988) developed a series of TAT-type cards to study the presence of and dynamics involved in child abuse. Three sets of cards are available: (a) the basic set of 25 cards depicting scenes pulling for possible child abuse; (b) a 10-card set for neglect; and (c) 5 cards to assess attitudes toward different courtroom themes. The Family Apperception Test, composed of 21 pictures of family interactions, is designed to assess family dynamics (Julian, Sotile, Henry, & Sotile, 1991). The Blacky Pictures Test (G. Blum, 1950, 1962, 1968) is another thematic-type test that is closely aligned to psychoanalytic theory. It presents children with pictures of a dog, Blacky, that is involved in situations consistent with psychoanalytic theory, such as themes surrounding oral, anal, and phallic stages of development.

Ritzler, Sharkey, and Chudy (1980) have criticized the TAT for producing negative, low-energy stories and for containing outdated pictures that are difficult for persons to identify with. To counter this, they developed the Southern Mississippi TAT (SM-TAT) using pictures derived from the *Family of Man* (Steichen, 1955) photo collection. They report that using the SM-TAT pictures produces stories with more activity, greater emotional tone, and relatively few variations in thematic content (Sharkey & Ritzler, 1985). More importantly, the results derived from the SM-TAT were more effective in discriminating different pathological groups than the TAT. Depressives produced gloomy stories, and psychotics demonstrated more perceptual distortions when compared with normals. A more recent but similar attempt is the eight-card Apperceptive Personality Test (APT), which has the advantages of an objective scoring system, a set sequence of card presentations, multiethnic pictures, and initial positive validity outcomes (Holmstrom, Karp, & Silber, 1994; Karp, Holmstrom, & Silber, 1989; Karp, Silber, Holmstrom, Banks, & Karp, 1992). Although the SM-TAT and APT are more modern, are based on a more rigorous methodology, and demonstrate greater diagnostic validity, the long tradition and extensive research associated with the TAT may make it difficult to supplant, even with potentially better instruments.

In addition to the TAT's derivatives, a number of different approaches to scoring and interpreting the TAT itself have been developed. The original approach by Murray involves assessing which character in the story is the "hero" or focal figure, and then quantifying the relative intensity of each expressed need on a five-point scale. Murray also includes measuring the forces of the hero/heroine's environment (press), types of outcomes, basic themes (themas), and interests and sentiments of the hero/heroine. In addition to Murray's system, many variations have been developed by authors such as Arnold (1962), Bellak (1975, 1986, 1993; Bellak & Abrams, 1997), Chusmir (1985),

Dana (1955), Eron (1950), McClelland (1971), Sokolowski, Schmalt, Langens, and Puca (2000), A. Thomas and Dudek (1985), Weston (1995), and Wyatt (1947). The extensive diversity of different systems led Murstein, in his 1963 review of the TAT, to remark: "There would seem to be as many thematic scoring systems as there were hairs in the beard of Rasputin" (p. 23). The most frequently used and updated system has been Bellak's (1975, 1986, 1993; Bellak & Abrams, 1997). His book on the TAT (*The TAT, CAT, and SAT in Clinical Use*) has undergone six editions and is perhaps the simplest and most frequently used of the available systems (Rossini & Moretti, 1997). As a result, his scoring method and interpretive approach have been included in this chapter.

Although the TAT is still relatively popular, there are signs that its use is decreasing for a variety of reasons (C. Piotrowski, 1999). Managed care organizations are increasingly requiring clinicians to use brief, focused, time-efficient and cost-effective instruments (see Chapter 13; Groth-Marnat, 1999, 2000b). In contrast, the TAT takes considerable time to administer, score, and interpret. Instruments that are far more focused on, and more empirically connected to, treatment planning have also been developed (see Chapter 14). In addition, professional psychologists are being expected to be proficient in an increasingly wider variety of instruments (i.e., neuropsychological assessment, behavioral assessment), leaving less time for the time-consuming training required for projectives. Alternative instruments, such as the rapid proliferation of rating scales during the last half of the 1990s, have helped to supplant the TAT/CAT's use with children and adolescents (Kamphaus et al., 2000). Finally, controversy regarding the psychometric properties of projectives has increased over the past few years (Garb, 1998b, 1999). Despite these trends, there are a number of noteworthy developments relevant to the TAT and its future. A number of more recent scoring systems that have made useful and valid predictions have been researched (see Cramer, 1999). For example, Weston et al.'s (1995) Social Cognition and Object Relations Scale has predicted relevant dimensions of personality disorders (S. Ackerman, Clemence, Weatherill, & Hilsenroth, 1999). In general, the object relations school has placed considerable interest in extending the knowledge and usefulness of the TAT. There has also been considerable interest in better understanding client narratives as they relate to therapeutic techniques, health outcomes, and the structure of personality (see Special Series in the *Journal of Clinical Psychology,* 1999, pp. 1175–1270). This has also helped to extend and better understand the meanings of TAT narratives (i.e., Cramer, 1996; Pennebaker & King, 1999). One possible future for the TAT (and its derivatives) might be that a combination of voice recognition technology and rapid computer analysis and interpretation of word clusters and sequences may make it a more practical, time-efficient technique, as well as allow the most validated interpretations to be selected and used.

THEORETICAL PERSPECTIVES

The TAT was originally developed based on Murray's concepts of personality. At the core of his concepts was a focus on how individuals interact with their environments—how people are affected by external forces and how their unique sets of needs, attitudes, and values influence their reaction to the world around them. Perhaps more than any other theorist, he has analyzed and clarified the concept of *needs.* The TAT itself

was originally conceptualized as a means of measuring the strengths of various needs as expressed by the designated hero in the story. A need can be either provoked by internal processes or, more frequently, the result of specific environmental forces. Murray developed a list of 28 needs that helped to specify the total possible needs that might be expressed in an individual's life (or reflected in TAT stories). To balance and complement the presence of needs, Murray also developed a list of possible forces in a person's environment, which he termed *press.* A total of 24 press were identified, and the relative strength of these press could also be scored on the TAT. As a result, Murray's theory is oftentimes referred to as *needs-press theory.*

To conceptualize units of behavior that result from the interaction between needs and press, Murray developed the term *thema.* A thema is a small unit of behavior that can combine with other thema to form *serial thema.* An individual's *unity thema* is the pattern of related needs and press that gives meaning to the largest portion of his or her behavior. For example, a core and overriding feature of an individual might be rebelliousness or martyrdom. This may be sufficiently well organized and powerful to override even primary (biological) needs, as amply demonstrated in the case of a martyr who is willing to die for his or her beliefs. A unity thema is derived from early infantile experiences and, once developed, repeats itself in many forms during an individual's later life. It operates largely as an unconscious force, and Murray (1938) described it as "a compound of interrelated-collaborating or conflicting-dominant needs that are linked to press to which the individual was exposed on one or more particular occasions, gratifying or traumatic, in early childhood" (pp. 604–605). The TAT was designed to assess both small units of thema and the larger, core aspects of an individual's unity themas.

Murray's theories of personality were obviously the main influence on the early development and use of the TAT. However, psychoanalytic, object relations, and theories of understanding narratives have also had a significant influence on conceptualizing, scoring, and interpreting the TAT. Psychoanalytic conceptualizations easily lend themselves to the interpretation of TAT stories. The cards themselves depict many images that are highly relevant to a psychodynamic perspective such as possible superego conflicts in Card 1 (boy sitting at a table with a violin in front of him) or castration anxiety related to Card 8BM (a boy in the foreground staring into space with two men in the background performing surgery on a patient lying down). Stories often relate to internal conflict and how the person deals with this conflict. The Bellak scoring system (Bellak & Abrams, 1997) is organized around classic psychodynamic domains such as the client's conceptualization of parental figures, main defenses against conflicts and fears, adequacy of the superego, and the main drives of the hero. Cramer's (1996) scoring system specifically includes scoring for the defense mechanisms of denial, projection, and identification.

TAT stories can also be understood as depicting the quality of a client's object relations. This is reflected in Weston's coding system (Social Cognition and Object Relations Scale; Weston, 1995) that focuses on understanding the following crucial areas of psychological functioning: (a) the client's internal representation of significant others, (b) quality of affect in relationships, (c) capacity for emotional investments and moral integrity, and (d) understanding the extent to which a person can understand interpersonal motivation. The themes surrounding TAT stories can be seen as ideal sources of

data to extract and elaborate on these areas of client functioning. As would be expected, researchers have used the TAT to better understand those disorders that lend themselves to poor object relations, such as narcissistic, borderline, and antisocial personalities (Ackerman et al., 1999; Cramer, 1999).

One of the core activities of being human seems to be the importance of constructing a coherent story. Much of therapy can be considered an interaction that helps the client to recreate the story of his or her life in a way that he or she can more easily live with. Thus, narrative can both reflect a person's current condition and be used as a means of creating change (Burns, 2001; Groth-Marnat, 1992). Certain patterns of narrative (reflecting on self versus describing external experiences) have been associated with better therapeutic outcomes (Angus, Levitt, & Hardtke, 1999). Research has also found that the word patterns that people use over time are quite reliable, and reflect not only how people cognitively organize their world, but also relevant dimensions of personality (Pennebaker & King, 1999). Having people write emotionally laden stories for as little as 15 minutes a day for four consecutive days has been found to result in less illness, positive increases in markers of immune system, higher grades, and lower reported pain levels (Pennebaker & King, 1999). It also seems that after people have been able to adequately understand how and why an event occurred, they are more able to deal with it the next time it occurs and can also move beyond it. This area of research has had a direct impact on the TAT in the form of new coding systems, greater understanding to how everyday narrative relates to TAT stories, and the factor structure of word patterns on the TAT (Pennebaker & King, 1999).

RELIABILITY AND VALIDITY

A subject's responses to the TAT involve complex, meaningful verbal material. Because of the complexity of this material, exact quantitative analysis is difficult, and interpretations are typically based more on a qualitative than a quantitative analysis of story content. Most methods of determining reliability, therefore, become problematic. This issue is further complicated because there are so many different scoring systems. Reliability (and validity) for one system may not mean that adequate reliability will be present for another system. However, some success in achieving adequate interscorer reliability has resulted from the development of quantitative scoring strategies and rating scales. This is especially true for the work of McClelland (1961) and Atkinson and Feather (1966), who developed complex scoring schemes for achievement, affiliation, and power. Interscorer reliability across different scoring systems has generally been found to be good, ranging between .37 and .90, with most reports .85 or higher (Pennebaker & King, 1999; Winter, 1999). However, although scorers can agree on the quantitative values assigned to different variables, these still constitute raw data and not conclusions. In other words, it remains questionable whether clinicians make the same inferences regarding personality based on the quantitative scores. Good interscorer reliability relating to areas such as the weighting of different needs has been achieved, but agreement between the conclusions based on these scores has typically not been adequately demonstrated. A further complication is the fact that, in actual practice, clinicians rely primarily on intuitive clinical judgment, use different sets of

instructions, and vary the number, type, and sequence of cards from one client to the next (Gieser & Stein, 1999; Karon, 2000). Thus, reliability in clinical contexts is likely to be considerably lower than under experimental conditions.

Another difficulty in determining reliability lies in the wide variability among different stories. If test evaluators wish to determine the internal consistency of the TAT, they are confronted with the fact that the various cards are not comparable (Cramer, 1999; Entwisle, 1972). They were designed to measure separate areas of a person's functioning. Thus, a strategy such as split half reliability is inappropriate. Not only are different stories in the same administration likely to be different, but so are the stories between two different administrations. Thus, measures of internal consistency have been (and would be expected to be) low (Entwisle, 1972). Likewise, when subjects were asked to tell different stories on different administrations, the test-retest reliabilities derived from quantitative scorings of various needs were low (Lindzey & Herman, 1955). In contrast, Lundy (1985) found that when subjects were requested to tell a similar story between one administration and the next, test-retest reliabilities achieved a respectable .56 (need for affiliation) and .48 (need for achievement). This suggests that the test-retest reliability of the TAT might be underestimated. However, the higher reliabilities found by Lundy might reflect merely the quality of memory rather than the stability of personality indices on the TAT.

Reviews of the TAT's validity have shown wide variability. Proponents of the test describe "impressive" and "strong" relationships, whereas critics have said that validity is "almost nonexistent." This disparity can be partially accounted for by differing interpretations of the data. One reviewer might be impressed by a correlation of .25 while another sees it as highly deficient. It would seem that not only is the TAT a projective test itself, but the research done on it likewise allows readers to project their biases, needs, and expectations onto the TAT. One factor that might help to explain the differences in results between studies is that the TAT has been found to be quite sensitive to the effects of instructions. Lundy (1988), for example, found that under conditions that were nonthreatening, neutral, and unstructured, there were moderate correlations between outside criterion measures and needs for achievement, affiliation, and power. When instructions were used that presented the TAT as a structured formal test or, especially, when any words were used that might have been interpreted as threatening (will "reveal imperfections" or "minor defects"), the correlations were nonsignificant. This suggests the interesting possibility that the wide variation in the findings of different studies may have been partially influenced by slight variations in instructions.

Studies attempting to determine criterion validity have shown a balance between positive and negative findings. One major problem lies in establishing agreed-on external criteria. If short-term overt behavior is used as the criterion, there is often little correspondence to test scores. For example, high aggression on TAT scores usually does not reflect the degree to which a person actually expresses aggressive behavior. However, it may still be valuable to understand a person's internal processes even though these are not outwardly expressed. When measures of needs on the TAT were compared with needs measured on tests such as the Personal Research Form, Edwards Personal Preference Schedule, Adjective Check List, and other forms of self-rated questionnaires, there was little correspondence (Megargee & Parker, 1968; Schultheiss & Brunstein, 2001; Spangler, 1992). These findings would seem to call into question the usefulness

of TAT protocols. However, a number of positive findings between the TAT and outside criteria have also been reported. An early and frequently cited study by Harrison (1940) found that diagnosis by a trained clinician using the TAT was accurate 75% of the time when assessing broad diagnostic categories. Similarly, a correlation of .78 was found when comparing TAT inferences with data from hospital records. The most extensively studied constructs have been achievement, affiliation, and power (Lundy, 1988; Spangler, 1992; Winter, 1999), and these, too, have had varying degrees of success when they were compared with outside criteria. Examples of positive results include a high need for achievement being associated with greater social attractiveness (Teevan, Diffenderfer, & Greenfield, 1986), need for affiliation being positively correlated with a preference for an internally directed orientation to tasks (Schroth, 1987), and need for achievement being positively related to grade point average (although this might have been confounded by verbal fluency; Lindgren, Moritsch, Thurlin, & Mich, 1986). In addition, Coche and Sillitti (1983) reported that the presence of depressive themes on the TAT was correlated with measures of depression on the MMPI and Beck Depression Inventory. Maitra (1983) found that the fantasies of highly effective executives differed significantly from those of relatively ineffective executives in that the more effective executives had more original themes, expressed a broader range of interests, were more intellectual, and could see beyond the individual details of their work.

In a meta-analysis of 105 studies, Spangler (1992) concluded that, on average, TAT correlations of behavior (.19 to .22) were slightly larger than correlations based on questionnaires (.13 to .15). Although these correlations were quite low, the TAT and questionnaires were relatively more effective in different situations. The TAT produced quite high correlations (.66) when subjects were required to spontaneously initiate (*internally*) their own behavior to achieve some activity (moderate task risk, task completion, response to time pressure). In contrast, self-report questionnaires were low in making these predictions, but relatively effective (.35) in predicting situations involving real-world behavior where social reinforcement (*external*) was present.

Like the studies on criterion validity, the work on construct validity has shown varying results. A representative confirmatory study provided support for the hypothesis that subjects who were experimentally frustrated produced subsequent stories in which the focal characters in the stories expressed increased aggression (Lindzey & Kalnins, 1958). Somewhat similarly, increased levels of stress have been associated with higher defense-related scores on the TAT (Cramer, 1998).

An important issue in the interpretation of criterion validity studies on the TAT and other projective devices is understanding the implications of different levels of interpretation. W. Klopfer (1983) summarized the earlier work of Leary (1957) by indicating that behavior can be based on outside observations (direct behavioral data, public communication), self-descriptions, or private symbolization. These three levels are often quite different from one another. For example, the observations by others are frequently quite discrepant from how a particular person perceives himself or herself. Likewise, a person's inner fantasy life (private symbolization) is often quite different from his or her public behavior. Projective tests such as the TAT primarily assess a person's inner life of private symbolization. McClelland, Koestner, and Weinberger (1989) have described the characteristics that the TAT measures as being *implicit* motives as opposed to more conscious self-attributions. Thus, the TAT measured motives and

underlying themes that are physiologically and nonconsciously connected to the person, and they influence long-term rather than short-term behaviors. This is supported in that TAT results do accurately predict long-term outcomes such as overall success in a person's career (Spangler, 1992). McClelland et al. (1989) even emphasize that it would be unreasonable for the TAT to correlate highly with immediate, conscious, short-term behavior because this is not what the TAT was designed to measure. Furthermore, it may not even be desirable for TAT data to relate to immediate external behavior because it is precisely this ability to access a person's inner life that makes projectives both unique and valuable. From a practical perspective, each clinician needs to evaluate these issues and establish the importance each places on having access to a person's inner world of private symbolization.

One argument against subjecting the TAT to strict psychometric scrutiny is that rigid objective studies do not represent the way in which the TAT is actually used in clinical practice (Karon, 2000). When experienced clinicians were requested to provide individual descriptions of persons based on TAT stories, the descriptions did tend to match independent descriptions based on case histories (Arnold, 1949; Harrison, 1940; Karon, 2000). However, although the descriptions by individual clinicians were fairly accurate, there was usually little agreement among different clinicians evaluating the same person. It might be argued that, because of the complexity and richness of the material, each clinician was tapping into different (but still potentially valid) aspects of the same person. The poor interrater reliability might then be interpreted as not representing inaccuracy, but rather different approaches to the material, with each of these approaches having potentially relevant meanings for the client being evaluated.

ASSETS AND LIMITATIONS

Despite questions related to the TAT's reliability and validity, its frequency of use over the past 40 years has remained essentially unchanged (Archer et al., 1991; Camara et al., 2000; Lubin et al., 1985). It is still rated in the top ten most frequently used instruments, and it has produced the fourth largest number of research studies (behind the MMPI, Wechsler intelligence scales, and Rorschach). One reviewer has summarized the incongruity between its popularity and its questionable validity by stating "there are still enthusiastic clinicians and doubting statisticians" (Adcock, 1965). In light of this controversy, it is especially important that clinicians fully understand the general assets and limitations involved with the TAT.

Like most projective techniques, it theoretically offers access to the covert and deeper structures of an individual's personality. There also may be less susceptibility to faking because the purpose of projective techniques is usually disguised, and the subject often slackens his or her conscious defenses while releasing unconscious material. However, because the TAT deals with verbally familiar material, there is a somewhat greater potential for the subject to bias and distort his or her responses when compared to the less familiar stimuli of the Rorschach. A further asset is the focus on the global nature of personality rather than on the objective measurement of specific traits or attitudes. These include not only emotional, motivational, interpersonal, and defensive characteristics but also general intellectual level, verbal fluency, originality,

and style of solving problems. Finally, there is ease of rapport. The TAT is generally regarded as intrinsically interesting and nonthreatening because there are no "wrong" answers and there are no direct questions related to personal and potentially sensitive information. However, certain types of individuals might still feel quite anxious and insecure with the TAT's relative lack of structure.

In contrast to these assets, the following general criticisms have been leveled at the TAT (and projective techniques in general). There has typically been difficulty establishing adequate internal consistency and test-retest reliability. Because inadequate normative data are generally lacking, clinicians often rely on clinical experience when they interpret the responses. The standardization in respect to administration and scoring is generally inadequate. Thus, the effectiveness of the technique is often more dependent on the clinician's individual skill than on the quality of the test itself. Most studies on the TAT, as well as the results coming from individual assessments, are confounded by verbal abilities, age, sex, intelligence, and reading ability (Klein, 1986). The TAT has also been shown to be quite sensitive to situational variables such as mood (McFarland, 1984), stress, sleep deprivation, and differences in instruction (Lundy, 1988). These variables can significantly alter test performance, thereby reducing the likelihood that stable aspects of personality are being measured. Finally, many of the validity studies on the TAT have been equivocal. In particular, several researchers have noted that there has been no increase in incremental validity when the TAT and most other projectives are used in a battery of tests (see Garb, 1984, 1998b; Klein, 1986; Lanyon & Goodstein, 1982).

In contrast to these limitations, one important asset is that the responses it produces from clients (verbal stories) are familiar rather than hidden and mysterious. Even a relatively untrained person can appreciate the differing themes, moods, and perspectives portrayed in the stories. The experienced clinician also profits from this inherent familiarity or approachability of the test data.

A further asset of the TAT is its origin within an academic-humanistic environment. It is not closely aligned with any particular school of thought and, therefore, can be approached from, and interpreted by, a number of different theoretical orientations. Furthermore, the TAT was developed from the study of normal individuals rather than by case studies or normative comparisons with disordered populations. This orientation has evolved directly out of Murray's belief that the proper beginning point for understanding personality is the intensive and detailed study of normal persons.

The TAT potentially provides a comprehensive evaluation of personality, which has sometimes been referred to as a *wide-band* approach. For example, among the comprehensive dimensions that the TAT can assess are these: a person's cognitive style, imaginative processes, family dynamics, inner adjustment, emotional reactivity, defensive structure, internal representation of significant people, general intelligence, and sexual adjustment (Bellak & Abrams; 1997; Gieser & Stein, 1999; W. Henry, 1956). The TAT also has some potential to evaluate areas such as creativity, level of affect, problem-solving skill, and verbal fluency. Thus, although the TAT is primarily concerned with providing insight into a person's fundamental needs and patterns of interaction, it can potentially give important information about a far wider range of areas. In particular, the TAT may bypass conscious resistance to provide themes that the person may not reveal upon direct questioning. For example, alcoholics who reported high levels of internal

locus of control on direct self-report questionnaires typically became highly externally oriented when locus of control was measured using a TAT-type instrument (R. Costello & Wicott, 1984). This might be interpreted as the TAT-type test bypassing their conscious denial and assessing a possibly more accurate, or at least different, level of private symbolization.

Although the TAT is potentially quite versatile, it is not self-sufficient. A number of authors have emphasized that the TAT yields optimal results only when included in a battery of tests (Anastasi & Urbina, 1997) and/or as a type of structured clinical interview (Obrzut & Cummings, 1983). In contrast, some reviewers have pointed out that the TAT and other projective tests do not help increase incremental validity in structured conditions and may even serve to reduce it (Garb, 1984, 1998b; Klein, 1986).

One unresolved dispute surrounding the TAT is the relationship between inner fantasies and overt behavior. It has been assumed by most projective test originators, including Murray, that fantasy productions can be used to predict covert motivational dispositions. However, it is questionable whether high fantasy productions in a certain area actually parallel overt behavior (Klinger, 1966; McClelland, 1966; Skolnick, 1966). In fact, fantasies may even serve to compensate for a lack of certain behaviors. It might be quite consistent for a highly repressed, overcontrolled person to have a high number of inner aggressive fantasies. In a 20-year longitudinal study of adolescents who obtained high TAT scores on need for achievement, the subjects were often *not* among those who subsequently showed upward social mobility. However, individuals who had shown upward social mobility typically obtained higher TAT need-for-achievement scores as adults. The interpretive significance here is that it might be better to see fantasy productions as samples of thoughts that may or may not accurately reflect overt behavior.

Practical difficulties associated with the TAT include the extensive amount of training required to properly learn the technique and its poor cost-effectiveness in terms of the time required for administration and scoring (Groth-Marnat, 1999; Haynes & Peltier, 1985). Simply obtaining biographical information, asking direct questions during an interview, or using rating forms or questionnaires might give similar information in a simpler and quicker manner. It is likely that any major developments in the TAT will occur through the use of computerized content analysis of natural language (Gottschalk, 2000). This may make using it time efficient, provide more versatility in scoring, and possibly help develop stronger psychometric properties.

A further liability is the subjectivity involved in both scoring and interpretation procedures. Although the various scoring methods have attempted to reduce the degree of subjectivity, intuitive judgment necessarily plays a significant role. This results in part from an inadequate development of norms, and from the fact that the norms that have been created are only a rough approximation of common story themes (see Typical Themes Elicited section). Frequently, in clinical practice, each person develops his or her own individual intuitive norms, based on experience. Thus, the clinician may have a general intuitive conception of what constitutes a "schizophrenic" or "narcissistic" story and will use this subjective schema during diagnostic or interpretive procedures. Reliance on clinical experience becomes indirectly encouraged both by the lack of precise normative data and, more importantly, by the belief that norms tend to decrease the richness and comprehensiveness of the material being studied. A possible

source of bias is that, because clinicians work predominantly with pathology, their firsthand experience of the characteristic reactions of normal people on tests such as the TAT is limited. This may result in their overemphasizing the pathological features of stories and experiencing difficulty when assessing fairly well-adjusted persons.

ADMINISTRATION

General Considerations

The TAT was intended to be administered in an interpersonal setting in which subjects verbally respond to pictures presented to them. However, when the examiner is absent, responses may be taped or written by the subject. The disadvantage of these latter procedures is that the subject's responses are often more contrived and clichéd, because more time is available to censor fantasy material.

The TAT materials consist of 20 cards on which ambiguous pictures are printed. The cards are numbered so that 20 cards can be presented to four different groups: adult males, adult females, boys, and girls. The back of each card is coded with a number and/or letters to designate which sex and/or age group the card is intended for. A number without a letter indicates the card is to be administered to all subjects regardless of age or sex. A number with "M" or "F" designates that the card is intended for males or females, and "B" or "G" designates boys or girls, respectively. There may also be a number and either BM or GF, indicating the card is to be given to boys/males or girls/females.

Although Murray originally recommended that all 20 cards be given, in actual practice shorter versions typically consisting of 8 to 12 selected cards are used (Bellak & Abrams, 1997; Haynes & Peltier, 1985; Karon, 2000). The selection of cards may be idiosyncratic to the patient's presenting problem or based on previous information derived from relevant history or other test data. For example, if depression and suicide are significant issues for the client, the examiner might administer cards 3BM, 13B, and 14 in an attempt to gather specific information regarding the dynamics of the client's condition. Specific cards may also be selected because they typically produce rich responses. Bellak and Abrams (1997) recommend that the following standard sequence of 10 cards be administered to both females and males in this exact order: 1, 2, 3BM, 4, 6BM, 7GF, 8BM, 9GF, 10, and 13MF. He further recommends that an essential sequence of cards to be administered to any males consists of: 1, 2, 3BM, 4, 6BM, 7BM, 11, 12M, and 13MF. The essential cards for any females are the following sequence: 1, 2, 3, 3BM, 4, 6GF, 7GF, 9GF, 11, and 13MF. If a reduced number of cards are used, it may be preferable to give the cards in the sequence numbered on the back.

For research purposes, a slightly different listing of card frequencies has been found by Keiser and Prather (1990), who reviewed 26 studies that specified which of Murray's cards were used. The 10 most frequent cards were 1, 2, 3BM, 3GF, 4, 5, 6BM, 6GF, 8BM, and 8GF.

During administration, the subject should be seated beside the examiner, with his or her chair turned away so that he or she cannot see the expressions on the examiner's face. Ideally, this creates a situation in which the subject is comfortable and relaxed, so that his or her imagination can freely respond to the cards. However, if some individuals do not feel comfortable when turned away from the examiner, they should be

allowed to sit in a position that is more relaxing for them. Of primary importance are establishing adequate rapport and keeping the subject comfortable and relaxed.

Instructions

Murray's original instructions from the TAT Manual (1943) are as follows:

> This is a test of imagination, one form of intelligence. I am going to show you some pictures, one at a time; and your task will be to make up as dramatic a story as you can for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. Speak your thoughts as they come to your mind. Do you understand? Since you have fifty minutes for ten pictures, you can devote about five minutes to each story. Here is the first picture. (p. 3)

This set of instructions is suitable for adolescents and adults of average intelligence and sophistication. However, the instructions should be modified for children, adults with minimal education or intelligence, and psychotics. For these types of individuals, Murray (1943) suggests that the examiner state:

> This is a story-telling test. I have some pictures here that I am going to show you, and for each picture I want you to make up a story. Tell what is happening before and what is happening now. Say what the people are feeling and thinking and how it will come out. You can make up any story you please. Do you understand? Well, then, here is the first picture. You have five minutes to make up a story. See how well you can do. (pp. 3–4)

Such instructions may, of course, be modified, elaborated, or repeated to meet the individual needs of each subject. Lundy (1988) recommends that the instructions be given in as neutral and nonthreatening a manner as possible, so that the person doesn't become defensive. Defensiveness is likely to compromise the quality and accuracy of information. Any references to the TAT as a "test" should be avoided. However, the instructions should clearly indicate that the client is to use some imagination and not merely provide a description of the pictures. Variations on the instructions should also emphasize the four requirements of the story structure:

1. Current situation.
2. Thoughts and feelings of the characters.
3. Preceding events.
4. Outcome.

The instructions, either in whole or in part, may be repeated at any time, particularly if the subject has given a story that is too short or too long, or if he or she has left out one or more of the four requirements. The TAT can potentially be useful for evaluating mentally retarded persons, but particular care needs to be taken to ensure that the instructions are concrete and explicit. The examiner may also want to check whether the instructions have been clearly understood. He or she may need to encourage the person at various times during the storytelling.

Procedure

Time

The time measured should begin when the picture is first presented and end when the subject begins his or her story. It is particularly important to notice any long pauses or hesitations. They may reflect a struggle with conflictual or anxiety-laden material.

Recording

A subject's complete responses should be recorded, along with any noteworthy behavioral observations: exclamations, stuttering, pauses, blushing, degree of involvement, and changes in voice inflection. Thus, the general purpose of recording is not only to develop a reproduction of the verbatim story content, but to assess how the person interacts with the picture. As mentioned previously, ongoing verbal involvement with the cards is the preferable form of administration. Having subjects write out their own stories allows time for critically evaluating and censoring their responses. There is no objection to the use of a tape recorder, although, under such conditions, it is helpful to have the examiner record noteworthy behavioral observations and obtain the clients' written consent.

Questioning and Inquiry

If a subject omits certain aspects of the story, such as the outcome or preceding events, the examiner should ask for additional information. This may take the form of questions, such as "What led up to it?" or "How does it end?" However, these requests for clarification or amplification should not be stated in such a way as to bias the stories or reveal the examiner's personal reaction. An optional, more detailed inquiry may be undertaken either after the entire administration of the cards, or directly after each story. Murray recommends that the inquiry occur only after the administration of all the cards. Sample inquiry questions may include: "What made you think of this story?" or "Do people you have mentioned in the story remind you of friends or acquaintances?" As with questioning, the inquiry should not be too forceful or it may produce defensiveness and withdrawal. The overall purpose of both the questioning and the inquiry is to produce an unhampered and free flow of the subject's fantasy material.

Order of Presentation

Usually, the cards should be administered according to their sequential numbering system. However, at times, the examiner may be interested in a specific problem and alter the sequence to more effectively obtain information concerning that problem area. For example, if the clinician is particularly interested in problems relating to family constellation in a male subject, the examiner might include some of the female series involving sisters, girlfriends, or wives.

Use of the TAT (or CAT) with Children

Instructions for children should, of course, be modified in accordance with their age and vocabulary. It is usually helpful to describe the test as an opportunity to tell stories or as an interesting game. In general, cards from the TAT should be based on the likelihood that children may easily identify with the characters. For use with children, the

TAT cards that have the highest number of interpretable responses and the lowest number of refusals are the following, in order of usefulness: 7GF, 18GF, 3GF, and 8GF. In contrast, the least helpful cards are 19, 18BM, 11, and 12BG (Bellak & Abrams, 1997).

The stories of children seem to be relatively easily influenced by recent events experienced via television, movies, and computer games. Children also tend to project their problems and conflicts into a story in a more direct and straightforward manner than adults. Often, there is little hidden meaning or masking of the relationships involved.

TYPICAL THEMES ELICITED

At the present time, no formal normative standards have been developed for the TAT. The "norms" that are available are descriptions of the typical themes that occur for the different cards combined with clinical experience with these themes. This knowledge should be accompanied by an awareness of possible significant variations from the more frequent plots. These can serve to alert the examiner to unique, and, therefore, more easily interpretable, types of stories. Deviations from clichéd or stereotyped responses can be significant in that they may represent important areas of conflict, creative thinking, or important features of the subject's overall personality. If the clinician is equipped with expectations regarding typical versus unusual responses, it will enable him or her to (a) observe more easily specific attitudes toward the central problem; (b) notice gaps where the inquiry can begin; (c) assess which type of information the subject resists, as indicated by the use of noncommittal clichés; and (d) notice any deviation from the expected information that may contain significant and interpretable responses.

Murray's TAT cards and Bellak's original version of the CAT are described and discussed next. The descriptions of each TAT card that follow are divided into three sections:

1. Brief description of card.
2. Plots frequently encountered.
3. A general discussion of the significance and overall usefulness of the card.

The descriptions of each TAT card are this author's characterization of the scene's content; the CAT descriptions are from Bellak and Abrams (1997, pp. 286–289). The discussion of each picture summarizes the work of Bellak and Abrams (1997), Murray (1943), and M. Stein (1981).

Thematic Apperception Test (TAT)

Picture 1

1. **Description of Card.** A boy is sitting at a table looking at a violin placed on the table in front of him.

2. **Frequent Plots.** Typical stories emerging from this card revolve around either a self-motivated boy who is daydreaming about becoming an outstanding violinist, or a

rebellious boy being forced by his parents, or some other significant authority figure, to play the violin.

3. **General Discussion.** This is often considered to be the most useful picture in the entire TAT (Bellak & Abrams, 1997). It usually elicits stories describing how the subject deals with the general issue of impulse versus control, or, in a wider sense, the conflict between personal demands and external controlling agents. It also aids in providing information about the client's relationship with his or her parents, by making it relatively easy to see whether the parents are viewed as domineering, controlling, indifferent, helpful, understanding, or protecting (Bellak & Abrams, 1997). This card frequently gives specific information regarding the need for achievement, and it is important to consider how any expressed achievement is accomplished.

Any variations from the frequent plots described should be taken into consideration. They are likely to provide important reflections of the subject's characteristic modes of functioning. For example, the attitude toward, and relationship with, any introduced figures, or their identification as parents or peers, should be given special attention. Also of importance are the way in which the issue of impulse versus control is handled, any themes of aggression that might emerge, and, particularly, the specific outcome of the story.

Picture 2

1. **Description of Card.** Country scene with a woman holding a book in the foreground. In the background, a man is working a field while a woman watches.

2. **Frequent Plots.** Stories for this card often involve a young girl who is leaving the farm to increase her education or to seek opportunities that her present home environment cannot provide. Usually, the family is seen as working hard to gain a living from the soil. The family values often center on maintaining the status quo.

3. **General Discussion.** This picture usually provides an excellent description of family relations. As with Picture 1, various themes relate to autonomy from the family versus compliance with the status quo. This is one of the only cards in the series that presents the subject with a group scene and thus gives information relating to how the individual deals with the challenge of people living together. The card itself deals with a younger woman and an older male and female. Thus, it elicits stories dealing with parent-child and heterosexual relationships. There is usually the added dimension of contrasting the new and the old, and demonstrating attitudes toward personal mobility and ambition. This card may elicit stories relating to competition by the younger daughter for the attention of both or one of the parents. In these stories, her rival is either a sibling, particularly an older female, or the other parent. The extent to which separations or alliances occur among the three figures represented can also be quite revealing. For example, the two women may be united against the male who is "merely a hired hand," or the older male and female may be united against the younger female. Within either of these possible formations, it is important to note the attributes of each person, and the patterns and styles of interaction. Because this card is relatively complex and has a large number of details, compulsive patients often spend an inordinate amount of time commenting and elaborating on the many small details.

Picture 3BM

1. **Description of Card.** A boy is huddled next to a couch. On the floor next to him is an ambiguous object that could be a set of keys or a revolver.

2. **Frequent Plots.** The stories usually center on an individual who has been emotionally involved with another person or who is feeling guilty over some past behavior he has committed. Drug abusers often perceive the person in the picture as an addict and interpret the "revolver" as a hypodermic needle.

3. **General Discussion.** This has been identified as one of the most useful pictures (Bellak & Abrams, 1997; Keiser & Prather, 1990) because it concerns themes of guilt, depression, aggression, and impulse control. The manner in which the object on the left is seen and described often gives good information regarding problems concerning aggression. For example, if the object is described as a gun, is it used or intended to be used for intra-aggression (the subject is going to use it to do damage to self) or for extra-aggression (the subject has used it, or is going to use it, to harm another person)? If it is used for externally directed aggression, what are the consequences, if any, for the focal figure as portrayed in the outcome? This picture is particularly important for depressed patients, whether male or female, because it can reveal important dynamics regarding the manner in which the depression developed and how it is currently being maintained. For example, denial of aggressive conflict may be represented by completely overlooking the gun or rendering it harmless by depicting it as a toy pistol or a set of keys. On the other hand, excessive hesitation and detailed consideration of what the object might be could represent a compulsive defense surrounding conflictual aggressive feelings. Because this picture contains a lone figure, attitudes toward the isolated self are often aroused. The picture might be particularly useful for drug abusers because it frequently brings out themes and attitudes toward overdosing, drug use, mechanisms for coping, self-destructive tendencies, and extent of social supports.

Picture 3GF

1. **Description of Card.** A woman is standing next to an open door with one hand grabbing the side of the door and the other holding her downcast face.

2. **Frequent Plots.** As with Picture 3BM, the stories usually revolve around themes of interpersonal loss and contemplated harm directed internally because of guilt over past behavior.

3. **General Discussion.** The same general trends that hold for Picture 3BM are also true here, in that both pictures tend to bring out depressive feelings. Frequently, however, Picture 3BM brings out somewhat richer stories and allows both males and females to identify easily with the central figure.

Picture 4

1. **Description of Card.** A woman is grabbing the shoulders of a man who is turning away from her.

2. **Frequent Plots.** The primary task is to form some conceptualization as to why the woman is restraining the man. Often, the woman is seen as an advice-giving moral agent who is struggling with the more impulsive and irrational man. In approximately

half the stories, the vague image of a woman in the background is brought into the story plot.

3. **General Discussion.** This picture typically elicits a good deal of information relating to the feelings and attitudes surrounding male-female relationships. Frequently, themes of infidelity and betrayal emerge, and details regarding the male attitude toward the role of women may be discussed. For example, the woman may be seen as a protector who attempts to prevent the man from becoming involved in self-destructive behavior, or as a siren who tries to detain and control him for evil purposes. Likewise, a woman's attitude toward past male aggressiveness and impulsiveness may be revealed.

A further area of interest is the vague image of a seminude woman in the background. This often provokes themes of triangular jealousy in which one or more characters have been betrayed. When this picture is described, it is important to note whether the woman is depicted as a sexually threatening person or is seen as more benign.

Picture 5

1. **Description of Card.** A woman is looking into a room from the threshold of a door.

2. **Frequent Plots.** In the most frequent plot, a mother has either caught her child misbehaving or is surprised by an intruder entering her house.

3. **General Discussion.** This picture often reveals information surrounding attitudes about the subject's mother in her role of observing and possibly judging behavior. It is important to note how the woman is perceived and how the situation is resolved. Is she understanding and sympathetic? Does she attempt to invoke guilt? Or, is she seen as severely restricting the child's autonomy? Sometimes, voyeuristic themes are discussed, including feelings related to the act of observing others' misbehavior. The examiner should note whether these feelings include guilt, anger, indifference, or fear, and the manner in which these feelings are resolved. Often, this card elicits paranoid fears of attack or intrusion by an outsider, represented by stories in which the woman is surprised by a burglar.

Picture 6BM

1. **Description of Card.** An elderly woman is standing parallel to a window. Behind her is a younger man with his face down. He is holding onto his hat.

2. **Frequent Plots.** This picture typically elicits stories of a son who is either presenting sad news to his mother, or attempting to prepare her for his departure to some distant location.

3. **General Discussion.** This picture can be important to include when testing males. It usually proves to be a rich source of information regarding attitudes and feelings toward their mothers or maternal figures in general. Because the stories usually revolve around a young man striving for independence, the specific manner in which the subject depicts this struggle is important. Does the struggle include an exaggerated amount of guilt? Is there unexpressed or even overt anger toward the older woman? Or, does the young man succumb to the woman's wishes? Of equal importance is the

mother's reaction to her son's behavior. To what extent does she control him, and how? It is also of interest to note whether the subject accepts the traditional mother-son version, or whether he or she chooses to avoid discussing this relationship directly. If such an avoidance is evident, how are mother-son themes depicted in other cards that may have elicited discussions of this area (i.e., picture 1 or 5)?

Picture 6GF

1. **Description of Card.** A young woman sitting on the edge of a sofa looks back over her shoulder at an older man with a pipe in his mouth who seems to be addressing her.

2. **Frequent Plots.** The man is usually seen as proposing some sort of an activity to the woman, and the plot often includes her reaction to this suggestion.

3. **General Discussion.** This card was originally intended to be the female counterpart to Picture 6BM, and it was hoped that it, too, would elicit attitudes and feelings toward paternal figures. However, because the two figures are often seen as being about equal in age, the card frequently does not accomplish its intended purpose. When clear father-daughter plots are not discussed, the picture reflects the subject's style and approach to unstructured heterosexual relationships. For example, the subject may describe the woman as being startled or embarrassed or, on the other hand, may have her respond in a spontaneous and comfortable manner. It is important to note the manner in which the man is perceived by the woman. Is he seen as a seducer? Does he offer her helpful advice? Is he intrusive? Or, is he perceived as a welcome addition? A person who mistrusts interpersonal relationships typically creates a story in which the man is intrusive and the woman's reaction is one of defensiveness and surprise. Subjects who are more trusting and comfortable usually develop themes in which the woman responds in a more accepting and flexible manner.

Picture 7BM

1. **Description of Card.** An older man is looking at a younger man, who appears to be peering into space.

2. **Frequent Plots.** Stories usually describe either a father-son relationship or a boss-employee situation. Regardless of which of these variations is chosen, the older man is most frequently in the position of advising or instructing the younger one.

3. **General Discussion.** This card is extremely useful in obtaining information about authority figures and, more specifically, the subject's own father. The picture deals with hierarchical personal relationships and usually takes the form of an older, more experienced man interacting with a younger, less experienced one. Thus, the card can clearly show how the subject deals with external demands and attitudes toward authority.

Picture 7GF

1. **Description of Card.** A young girl is seated on a couch and is holding a doll in her hands. Behind her is an older woman who appears to be reading to her out of a book.

2. **Frequent Plots.** This picture is usually perceived as a mother and her daughter, with the mother advising, consoling, scolding, or instructing the child. Less frequently, there are themes in which the mother is reading to the child for pleasure or entertainment.

3. **General Discussion.** The intention here is to bring out the style and manner of mother-child interaction. When older women are the subjects, the picture often elicits feelings and attitudes toward children. Because both figures are looking away, either figure is sometimes perceived as rejecting the other. Thus, the card often elicits negative feelings and interactions, and it is important to note how these feelings are resolved, expressed, or avoided. Sometimes the older woman is described as reading a fairy story to the younger girl. Often, the most instructive data then comes from the fairy story itself.

Picture 8BM

1. **Description of Card.** A young boy in the foreground is staring directly out of the picture. In the background is a hazy image of two men performing surgery on a patient who is lying down.

2. **Frequent Plots.** Stories revolve around either ambition (the young man may have aspirations toward becoming a doctor) or aggression. Frequently, the aggressive stories relate to fears of becoming harmed or mutilated while in a passive state. Another somewhat less frequent theme describes a scene in which someone was shot and is now being operated on.

3. **General Discussion.** The picture can be seen as a thinly veiled depiction of a young man's oedipal conflicts, with concomitant feelings of castration anxiety and hostility. Thus, it is important to note what feelings the boy or other characters in the story have toward the older man performing the surgery. If the story depicts a need for achievement expressed by the younger man, it is also likely that he will identify with the older one and perhaps use him as an example. If this is the case, the details of how the identification takes place and specific feelings regarding the identification may be helpful.

Picture 8GF

1. **Description of Card.** A woman is sitting on a chair staring into space with her chin resting in her hand.

2. **Frequent Plots.** Because this picture is vague and nonspecific, extremely diverse plots are developed and there are no frequently encountered themes.

3. **General Discussion.** This picture is difficult to generalize about. Typically, it produces somewhat shallow stories of a contemplative nature.

Picture 9BM

1. **Description of Card.** Four men in a field are lying against one another.

2. **Frequent Plots.** Stories typically provide some explanation of why the men are there and frequently describe them either as homeless wanderers or as working men who are taking a much-needed rest.

3. **General Discussion.** This picture is particularly helpful in providing information about relations with members of the same sex. Are the men comfortable with one another? Is there any competitiveness? Is the central person in the story merely observing the four men, or is he one of the four men in the picture? Sometimes, homosexual tendencies or fears regarding such tendencies become evident in the story plot. Social prejudice surrounding attitudes toward "lazy," lower class, or unemployed persons often becomes apparent, particularly when the men in the picture are seen as homeless.

Picture 9GF

1. **Description of Card.** A woman in the foreground is standing behind a tree, observing another woman who is running along a beach below.

2. **Frequent Plots.** Usually, the two women are seen as being in some sort of conflict, often over a man. Frequently, either in addition to this theme or in a separate story, the woman "hiding behind" the tree has done something wrong. It is very unusual to have a story in which cooperation between the women is the central plot.

3. **General Discussion.** This card basically deals with female peer relations and is important in elaborating on issues such as conflict resolution, jealousy, sibling rivalry, and competitiveness. Because the figure standing behind the tree is carefully observing the woman on the beach, stories may provide details surrounding paranoid ideation. At the very least, the dynamics of suspiciousness and distrust are usually discussed. Frequently, a man is introduced into the story, often in the role of a long lost lover whom one or both of the women are running to meet, or a sexual attacker, from whom the woman on the beach is attempting to escape.

Picture 10

1. **Description of Card.** One person is holding his or her head against another person's shoulder. The gender of the two persons is not defined.

2. **Frequent Plots.** Stories usually center around some interaction between a male and a female, and may involve either a greeting between the two or a departure.

3. **General Discussion.** This card often gives useful information regarding how the subject perceives male-female relationships, particularly those involving some degree of closeness and intimacy. It might be helpful to notice the relative degree of comfort or discomfort evoked by emotional closeness. A story of departure or of termination of the relationship may be reflective of either overt or denied hostility on the part of the subject. Sometimes, males interpret the embrace as involving two males, which may suggest the possibility of a repressed or overt homosexual orientation.

Picture 11

1. **Description of Card.** On a road in a chasm, several figures are proceeding along a path toward a bridge. Above them and against the side of a cliff appears to be a dragon.

2. **Frequent Plot.** Typically, stories of attack and escape are elicited in which the subject takes into account the dragon, the path, and the obscure figures in the distance.

3. **General Discussion.** Because the form of this picture is quite vague and ambiguous, it is a good test of the subjects' imaginative abilities and their skills in integrating irregular and poorly defined stimuli. The picture also represents unknown and threatening forces, and reflects the manner in which the subjects deal with fear of attack. Thus, the examiner should take note of whether the characters in the story escape or instead become victims of their attackers. If they escape, how effective and coherent was the plan they devised to avoid danger? Were they instead saved by chance or "the forces of fate"? Subjects' stories can often suggest the degree to which they experience a sense of control over their environment and the course of their lives.

The dragon may be seen as coming out of the cliff and attacking people (representing aggressive forces in the environment), or as a protecting creature whom the characters are using for refuge and safety (a need for protection). Such themes can suggest aspects of the subjects' internal framework and mood. For example, subjects reporting stories of "everything being dead," suggests a depressive, impoverished inner state.

Picture 12M

1. **Description of Card.** A man with his hand raised is standing above a boy who is lying on a bed with his eyes closed.

2. **Frequent Plots.** Stories center on illness and/or the older man's use of hypnosis or some form of religious rite on the younger, reclining figure.

3. **General Discussion.** The picture often elicits themes regarding the relationship between an older (usually more authoritative) man and a younger one. This can be significant in predicting or assessing the current or future relationship between the therapist and the client. The manner in which the older man is perceived is particularly important. Is he sympathetic and giving aid, or is he described in more sinister terms? Thus, the picture can represent specifics of the transference relationship and, as such, can be an aid in interpreting and providing feedback to the client regarding this relationship. It can also be used to predict a client's attitude toward, and response to, hypnotic procedures. Stories related to this picture may also represent whether passivity is compatible with a subject's personality or is regarded with discomfort. In particular, subjects frequently reveal attitudes toward some external controlling force.

Picture 12F

1. **Description of Card.** A portrait of a woman is in the foreground; an older woman holding her chin is in the background.

2. **Frequent Plots.** Stories center on the relationship or specific communications between the two figures.

3. **General Discussion.** This picture elicits descriptions and conceptions of mother figures. The background figure is frequently seen as a mother-in-law who has a variety of evil qualities. Often, these negative qualities are feelings that the subject has toward her own mother but can indirectly, and, therefore, more safely, project onto the figure of a mother-in-law.

Picture 12BG

1. **Description of Card.** A country setting depicts a tree, with a rowboat pulled up next to it. No human figures are present.

2. **Frequent Plots.** Stories frequently center on themes of loneliness, peace, or enjoyment of nature.

3. **General Discussion.** With suicidal or depressed subjects, there may be an elaboration of feelings of abandonment and isolation—for example, someone has been lost or has fallen from the boat. More stable, adjusted subjects are likely to discuss the peace of being alone in the woods and perhaps of fishing or having gone fishing further down the stream.

Picture 13MF

1. **Description of Card.** A young man is standing in the foreground with his head in his arms. In the background is a woman lying in a bed.

2. **Frequent Plots.** The most frequent plot centers on guilt induced by illicit sexual activity. Themes involving the death of the woman on the bed and the resulting grief of the man, who is often depicted as her husband, are somewhat less frequent.

3. **General Discussion.** This picture is often considered to be helpful in revealing sexual conflicts. In a general way, it provides information on a subject's attitudes and feelings toward his or her partner, particularly attitudes just before and immediately following sexual intercourse. Stories in which there are overt expressions of aggression or revulsion are significant variations and should be noted as relatively unusual. In particular, the relation between a subject's aggressive and sexual feelings is frequently portrayed.

Because this picture has a relatively large number of details, obsessive-compulsive personalities frequently spend an excessive amount of time describing and explaining these details. This approach may be particularly evident when the picture has a shock effect and may, therefore, create anxiety. The obsessive-compulsive's style of handling anxiety by externally focusing on detail is then brought out.

Picture 13B

1. **Description of Card.** A boy is sitting in the doorway of a log cabin.

2. **Frequent Plots.** Themes of loneliness and stories of childhood are often elicited. However, because the stimulus is somewhat vague, the content and the nature of these stories tend to be extremely varied.

3. **General Discussion.** This picture may help both adults and children to reveal attitudes toward introspection or loneliness. In adults, it frequently elicits reveries involving childhood memories.

Picture 13G

1. **Description of Card.** A girl is climbing a flight of stairs.

2. **Frequent Plots.** The plots are similar to Picture 13B, usually involving themes of loneliness and/or distant childhood memories.

3. **General Discussion.** This picture lacks the specificity and impact found in other TAT cards. It usually produces stories that are highly varied but lacking in richness and detail. Like Picture 13B, it can sometimes be useful in depicting a subject's attitude toward loneliness and introspection.

Picture 14

1. **Description of Card.** A person is silhouetted against a window.
2. **Frequent Plots.** This card produces themes of contemplation, wish fulfillment, or depression, or feelings related to burglary.
3. **General Discussion.** If a subject's presenting problem is depression, especially if there is evidence of suicidal ideation, this card, along with Picture 3BM, is essential. This type of subject often describes the figure in the picture and, more importantly, discusses the events, feelings, and attitudes that led up to the current self-destructive behavior. It becomes important to investigate, during the inquiry phase of examination, the particular methods and styles of problem solving that the story character has attempted or is attempting. Also significant are the character's internal dialogues and personal reactions as he or she relates to different life stresses.

This picture may also reveal the subject's aesthetic interests and personal philosophical beliefs or wish fulfillments. If a story involving burglary is depicted, it can be useful to consider the character's level of impulse control and guilt, or the consequences of his or her behavior. For example, is the character apprehended and punished for his or her behavior, or is he or she allowed to go free and enjoy the profits of his or her misdeeds?

Picture 15

1. **Description of Card.** A man is standing among tombstones with his hands clasped together.
2. **Frequent Plots.** Themes usually revolve around beliefs or events surrounding death and a hereafter.
3. **General Discussion.** Stories from Picture 15 reflect the subject's particular beliefs about, and attitudes toward, death and the dying process. For example, death may be viewed as a passive, quiet process, or, in contrast, it can be experienced as a violent, aggressive situation. If the subject is having an extremely difficult time coping with the death of a friend or relative, the themes on Picture 15 can provide useful information as to why this difficulty is being experienced. For example, the story may reveal a method of adjustment based on excessive denial and a seeming inability to engage in grieving, from which a lack of resolution results. The story might also indicate unexpressed and problematic anger directed toward the dead person, because of a sense of abandonment.

Picture 16

1. **Description of Card.** Blank card.
2. **Frequent Plots.** Stories from this card are highly varied. It frequently elicits narratives related to a person's life (current marital, family, and personal situation) and, to a lesser extent, idyllic, defensive, catastrophic, and achievement-oriented concerns.

3. **General Discussion.** Instructions for this card are: Imagine a picture and then tell a story about it. From subjects with vivid and active imaginations, this card often elicits extremely rich, useful stories; and the amount of detail and complexity in a person's stories have been found to correlate with different measures of creativity (Wakefield, 1986). The card does little to shape or influence the subject's fantasy material and can thus be seen as a relatively pure product of his or her unconscious. However, for anxious, resistant, or noncreative subjects, this card is often of little or no value because the stories are usually brief and lack depth or richness. In considering the story, it is helpful to note whether the depiction involves a scene that is vital and optimistic, or one that is desolate and flat. This card's value can be increased by repeating the instructions, which stress that the person must provide a complete story (preceding events, current situation, and outcome), and giving the card as the last one in a series. Its value derives from both its total lack of structure and usefulness across different ages, ethnic backgrounds, and assessment goals.

Picture 17BM

1. **Description of Card.** A naked man is climbing up (or down) a rope.
2. **Frequent Plots.** Stories usually involve someone escaping from a dangerous situation or an athletic event of a competitive nature.
3. **General Discussion.** Because the card depicts a naked man, attitudes regarding the subject's personal body images are often revealed. They in turn may bring out themes of achievement, physical prowess, adulation, and narcissism. Possible homosexual feelings or anxiety related to homosexuality also becomes evident in the stories of some subjects.

Picture 17GF

1. **Description of Card.** A female is standing on a bridge over water. Above the bridge is a tall building, and behind the building the sun is shining from behind clouds.
2. **Frequent Plots.** A great variety of stories are elicited, although themes surrounding departure and social or emotional distance do occur with some frequency.
3. **General Discussion.** Attitudes toward a recent separation or the impending arrival of a loved one are sometimes described. This card can be particularly useful in cases of suicidal depression, where the figure on the bridge is perceived as contemplating jumping off, as a last attempt to resolve her difficulties. As with Pictures 3BM and 14, an inquiry into the specific difficulties the story character has encountered and the manner in which she has attempted to resolve these difficulties can often reflect the subject's manner and style of coping with his or her own difficulties. Personal reactions to, and internal dialogue involving, life stresses can also be extremely informative. However, some of this material is available only through a more detailed inquiry, after the initial story has been given.

Picture 18BM

1. **Description of Card.** A man dressed in a long coat is being grabbed from behind. Three hands are visible.

2. **Frequent Plots.** Typical themes involve either drunkenness on the part of the figure who is being supported by the three hands, or stories in which he is being attacked from behind.

3. **General Discussion.** This picture, more than any of the others, is likely to produce anxiety because of the suggestive depiction of invisible forces attacking the figure. Thus, it is important to note how the subject handles his or her own anxiety, as well as how the story character deals with his situation. Does the latter see himself as the victim of circumstance in which he is completely helpless? If so, how does he eventually resolve his feelings of helplessness? Is the helplessness a momentary phenomenon, or is it an ongoing personality trait? If the character is seen as the recipient of hard luck, then specifically what situation does the subject perceive as comprising hard luck? Exaggerated aggressiveness or attitudes toward addiction are also sometimes identified with this picture.

Picture 18GF

1. **Description of Card.** A woman has her hands around the throat of another woman. In the background is a flight of stairs.

2. **Frequent Plots.** Aggressive mother-daughter interactions or sibling relationships are often disclosed in response to this picture.

3. **General Discussion.** The manner in which the subject handles aggressive, hostile relationships with other women is the primary type of information this picture elicits. Particular note should be made of what types of events trigger this aggressiveness, and of the manner in which the conflict is or is not resolved. Does the character submit passively, withdraw from the relationship, plot revenge, or negotiate change? Feelings of inferiority, jealousy, and response to being dominated are also often described. Although the representation of aggressiveness in the picture is quite explicit, subjects occasionally attempt to deny or avoid this aggressiveness by creating a story in which one figure is attempting to help the other one up the stairs. This may point to general denial and repression of hostility on the part of the subject.

Picture 19

1. **Description of Card.** A surreal depiction of clouds and a home covered with snow.

2. **Frequent Plots.** Stories are highly varied because of the unstructured and ambiguous nature of the stimuli.

3. **General Discussion.** Because this is one of the more unstructured cards, the subject's ability to integrate disparate visual stimuli is tested. For certain subjects, the ambiguous nature of this picture can create anxiety and insecurity. The examiner can then observe how the subject handles his or her anxiety in the context of the story. Often, the stories produced deal with impersonal aggression from forces such as nature or the supernatural.

Picture 20

1. **Description of Card.** A hazy, nighttime picture of a man leaning against a lamppost.

2. **Frequent Plots.** Stories range from the benign theme of a late evening date to more sinister circumstances, perhaps involving a gangster who is in imminent danger.

3. **General Discussion.** The picture often elicits information regarding a subject's attitude toward loneliness, darkness, and uncertainty. Fears may be stated explicitly through gangster stories. As with Picture 18BM, the method of handling these fears and the examinee's response to physical danger should be noted.

Children's Apperception Test (CAT)

The following descriptions of, and typical responses to, pictures on the CAT are adapted from Bellak and Abrams (1997, pp. 286–289).

Picture 1

1. **Bellak's Description.** Chicks seated around a table on which is a large bowl of food. Off to one side is a large chicken, dimly outlined.

2. **Discussion.** Stories typically revolve around concerns relating to eating or sibling rivalry. The sibling rivalry may center on who is the best behaved, what the consequences of this behavior are, and which one gets more to eat. To obtain useful information on this card, it is particularly important to decide which character the subject identifies with. Food may be seen as reward for "good" behavior, or, conversely, when withheld, as punishment for "bad" behavior.

Picture 2

1. **Bellak's Description.** One bear is pulling a rope on one side, while another bear and a baby bear pull on the other side.

2. **Discussion.** Of particular importance in interpreting this picture is whether the bear who is helping the baby bear is seen as a male (father figure) or a female (mother figure). The struggle depicted can be seen either as a playful game of tug-of-war, or as a struggle involving a high degree of seriousness and aggression. For example, the loser(s) may end up falling off the edge of the rock and into a pool of dangerous animals. In the most recent revision of the CAT, the large bears were made equal in size, to avoid having the largest bear (previously depicted on the right) identified as the father.

Picture 3

1. **Bellak's Description.** A lion, with pipe and cane, sits in a chair; in the lower right corner, a little mouse appears in a hole.

2. **Discussion.** Because the lion is pictured with the characteristic symbols of authority (pipe and cane), this picture elicits attitudes and feelings toward father figures. It is important to note whether this figure is seen as benevolent and protecting, or dangerous and threatening. Sometimes, the subject defensively attempts to minimize the threat of the lion by reducing him to a helpless cripple who needs a cane just to move around.

Most children notice the mouse in the hole and blend it into their stories. Because the mouse and the lion are frequently seen in adversary roles, it is important to note

how the threatening presence of the lion is handled. Is the mouse completely under the control of the lion, and does it adapt by being submissive and placating? On the other hand, the mouse may be described as clever and manipulating, to trick and outsmart the lion. When subjects switch their identification back and forth between the lion and the mouse, some role confusion is suggested. This may be particularly true of enmeshed families or families in which the father is unable to set limits effectively.

Picture 4

1. **Bellak's Description.** A kangaroo who has a bonnet on her head is carrying a basket with a milk bottle. In her pouch is a baby kangaroo with a balloon; on a bicycle, there is a larger kangaroo child.

2. **Discussion.** As in Picture 1, this card elicits themes of sibling rivalry and, occasionally, themes revolving around a wish for regression, as demonstrated when the subject identifies with the baby kangaroo in the pouch. A regressive theme is particularly strong when a subject, who is in reality the oldest or middle child, identifies with the kangaroo in the pouch. On the other hand, a child who is actually the youngest may identify with the oldest kangaroo, thereby suggesting a strong need for autonomy and independence. On occasion, a theme of flight from danger may be introduced.

Picture 5

1. **Bellak's Description.** A darkened room contains a large bed in the background and a crib in the foreground in which there are two baby bears.

2. **Discussion.** Stories relating to attitudes and feelings about what occurs when parents are in bed are frequent responses to this card. They may involve aspects such as curiosity, conjecture, confusion, rejection, anger, and envy on the part of the children. Descriptions of the two bears in the foreground may also center on themes of sexual manipulation and mutual exploration.

Picture 6

1. **Bellak's Description.** A darkened cave shows two dimly outlined bear figures in the background and a baby bear lying in the foreground.

2. **Discussion.** This picture and Picture 5 elicit stories of parental bedtime activity. However, this picture tends to enlarge on and extend themes that have only begun to develop in Picture 5. Stories may also revolve around jealousy of the perceived intimacy between parents, or they may reflect possible feelings about masturbation on the part of the baby bear in the foreground.

Picture 7

1. **Bellak's Description.** A tiger with bared fangs and claws leaps at a monkey that is also leaping through the air.

2. **Discussion.** The subject often discusses his or her fears of aggression and characteristic manner of dealing with it. At times, the anxiety produced by this picture may

result in an unwillingness to respond to it at all. On the other hand, the subject's defenses may be either effective enough, or perhaps unrealistic enough, for him or her to transform the picture into a harmless story.

Picture 8

1. **Bellak's Description.** Two adult monkeys are sitting on a sofa drinking from tea cups. One adult monkey in the foreground is sitting on a hassock talking to a baby monkey.

2. **Discussion.** The subject often discusses his or her relative position and characteristic roles in the family. The description of the dominant monkey in the foreground as either a mother or a father figure should be noted as a possible indication of who has more control in the family. It is also significant to note how the dominant monkey is described. Is he or she threatening and controlling, or helpful and supportive?

Picture 9

1. **Bellak's Description.** A darkened room is seen through an open door from a lighted room. In the darkened one, there is a child's bed in which a rabbit sits up looking through the door.

2. **Discussion.** Typically, responses revolve around a subject's fears of darkness, possible desertion by parents, and curiosity as to what is occurring in the next room.

Picture 10

1. **Bellak's Description.** A baby dog is lying across the knees of an adult dog; both figures have a minimum of expressive features. The figures are set in the foreground of a bathroom.

2. **Discussion.** A child's attitudes and feelings about misbehavior and its resulting punishments are usually discussed in response to this card. In particular, his or her conceptions of right and wrong are often revealed. This picture is a good indicator of the child's degree of impulse control and his or her attitude toward authority figures when their role involves setting limits.

SCORING PROCEDURES

Since the publication of the original TAT Manual in 1943, there have been numerous methods of scoring and interpretation. Whenever a large number of different approaches are given to explain a particular phenomenon, it is usually a strong indication that none of them are fully adequate and all of them have significant shortcomings. This is true of the many alternate scoring and interpretation methods for the TAT. Difficulty arises primarily because of the type of information that is under investigation. Fantasy productions involve extremely rich and diverse information, which is difficult to place into exact and specific categories. Even the selection of which categories to use is open to question. For example, Murray prefers a listing and weighting of the primary needs

and press expressed in the stories, whereas Arnold (1962) emphasizes a restatement of the essential theme of the story on an interpretive level so as to highlight the basic meaning or moral. After deciding which method to use and evaluating the stories according to this method, the examiner is able to infer qualities of the subject's personality according to the categorization that is based on the specific method selected. Whether this final inference is valid and accurate is open to question and depends on a number of variables, including the skill and experience of the examiner, comparison with themes derived from other test data, and whether the state of the subject at the time of examination is representative of his or her usual orientation to the world.

For the purposes of this chapter, Bellak's (1954, 1993), and Bellak and Abram's (1997) method of interpretation is described. It is fairly comprehensive, easy to score, relatively concise, frequently updated, and the most frequently used of the various systems (Rossini & Moretti, 1997). Bellak's approach involves a certain degree of quantification in that interpreters are requested to rate the stories along several areas, according to the different story styles and contents. The goal is not so much to achieve a diagnosis of the subject as to obtain a description of how the subject confronts and deals with basic universal life situations. Each of the stories can be conceptualized as a series of common social situations depicting interpersonal relations. The manner in which the person constructs what he or she believes is occurring in these situations reveals a dominant pattern of social behavior as well as internal needs, attitudes, and values (Bellak & Abrams, 1997).

The specific scoring of the TAT and CAT cards can be organized on the Bellak TAT and CAT Analysis Sheet (see Table 11.1). The sheet provides a guide and frame of reference for TAT analysis that can be used later to organize and generate hypotheses about the person. It is intended to be used with a typical administration of 10 cards.

Using the long form of the scoring system, each one of the cards/stories is scored on a single Analysis Sheet. The overall story themes and contents can then be analyzed by noting the common themes and unique features throughout the different sheets. A shorter form is also available, consisting of using the Analysis Sheet as shown in Table 11.1, but simply having rows to the right of the 10 scoring categories to indicate the scoring for each card/story. Thus, 10 cards/stories might be scored in 10 consecutive rows to indicate Story No. 1, Story No. 2, and so on. At the end of this sequence is a Summary section, which the practitioner can use to organize conclusions. The summary categories are designated as follows:

1–3. Unconscious structure and drives of the subject (derived from scoring categories 1–3: Main Themes, Main Hero, and Main Needs and Drives of Hero).
4. Conception of world.
5. Relationship to others.
6. Significant conflicts.
7. Nature of anxieties.
8. Main defenses used.
9. Superego structure.
10. Integration and strength of ego.

Table 11.1 Bellak TAT and CAT analysis sheet

1. **Main Theme:** (diagnostic level: if descriptive and interpretive levels are desired, use a scratch sheet or page 5)
2. **Main hero:** age _____ sex _____ vocation _____ abilities _____
   interests _____ traits _____ body image _____
   adequacy (✓,✓,✓✓✓✓) and/or self-image _____
3. **Main needs and drives of hero:**
   a) behavioral needs of hero (as in story): _____
      implying: _____
   b) figures, objects, or circumstances *introduced:* _____
      implyng need for or to: _____
   c) figures, objects, or circumstanced *omitted:* _____
      implyng need for or to: _____
4. **Conception of environment (world) as:** _____
5. a) Parental figures (m _____ f _____) are seen as _____
      and subject's reaction to a is _____
   b) Contemp figures (m _____ f _____) are seen as _____
      and subject's reaction to b is _____
   c) Parental figures (m _____ f _____) are seen as _____
      and subject's reaction to c is _____
6. **Significant conflicts:** _____
   _____
7. **Nature of anxieties: (✓)**
   of physical harm and/or punishment _____
   of disapproval _____
   of lack or loss of love _____ of illness or injury _____
   of being deserted _____ of deprivation _____
   of being overpowered and helpless _____ lonely _____
   of being devoured _____ other _____
8. **Main defenses against conflicts and fears: (✓)**
   repression _____ reaction-formation _____ splitting _____
   regression _____ denial _____ introjection _____
   isolation _____undoing _____
   rationalization _____ other _____
9. **Adequacy of superego as manifested by "punishment" for "crime" being ( )**
   appropriate _____ inappropriate _____
   too severe (also indicated by immediacy of punishment) _____
   inconsistent _____ too lenient _____
   also: _____
   delayed initial response of pauses _____
   stammer _____ other manifestations of superego interference _____
10. **Integration of the ego, manifesting itself in (✓,✓✓,✓✓✓)**
    Hero: adequate _____ inadequate _____
    outcome: happy _____ unhappy _____
    realistic _____ unrealistic _____
    drive control _____
    thought processes as revealed by plot being : (✓,✓✓,✓✓✓)
    stereotyped _____ original _____ appropriate _____
    complete _____ incomplete _____ inappropriate _____
    syncretic _____ concrete _____ contaiminated _____

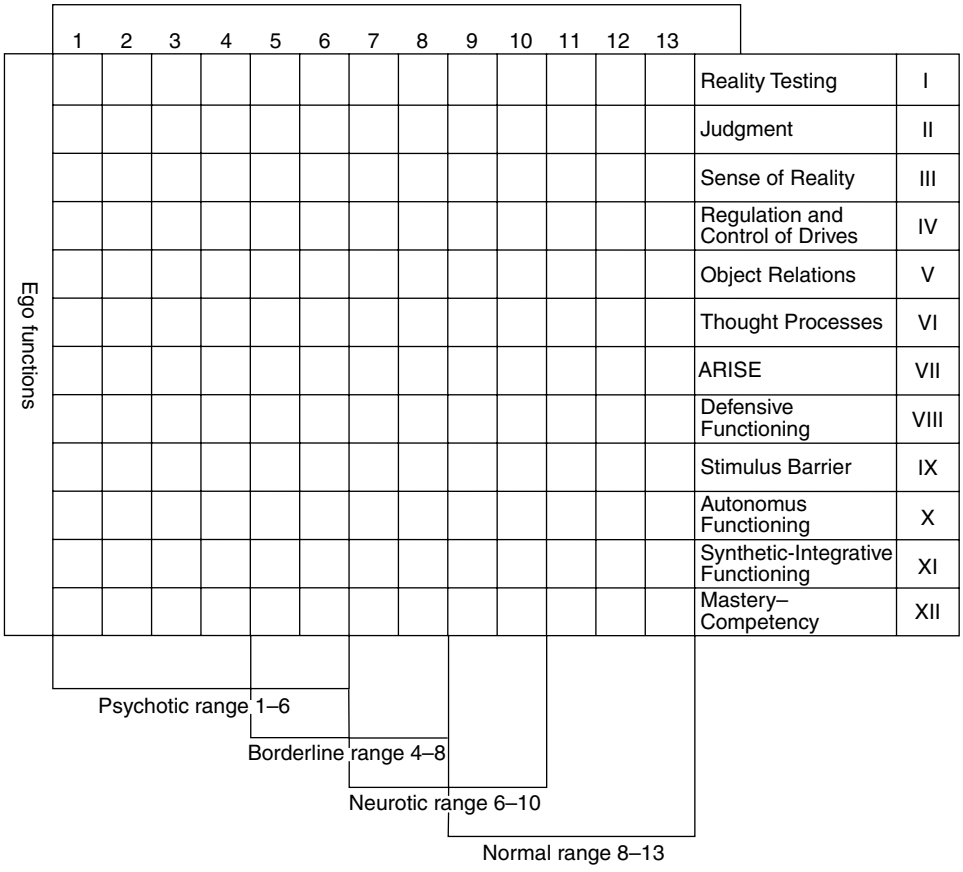
Intelligence _____
Maturation level _____
Organic Signs _____

Note: Reprinted with minor modification by permission of C.P.S. Inc., P.O. Box 83, Larchmont, NY 10538.

A table is also included (see Table 11.2), to provide a format for rating the person's ego functions. The combination of the summaries and ratings of ego functions serves as the actual interpretation of the TAT protocol. A short form is available from C.P.S., Inc. (see Table 11.2), and variations on the short form and long forms are available from The Psychological Corporation.

For each of the scoring categories, practitioners should abbreviate their observations about the person. In some sections, practitioners are asked to indicate the levels of importance or strength for the person by putting one check (✓—mere presence of characteristic), two checks (✓✓—moderate) or three checks (✓✓✓—strong). The entire scoring and interpretation procedure typically takes a half-hour. The ten scoring categories are described in the following sections. An attempt has been made to summarize and clarify as much as possible the descriptions provided by Bellak and Abrams (1997).

Table 11.2 Ego function assessment from TAT data



Source: From *Ego Functions in Schizophrenics, Neurotics, and Normals* by L. Bellak, M. Murvich, and H. Gediman, 1973, New York: John Wiley & Sons. Copyright 1973 by C.P.S., Inc.

1. Main Theme

This section requires the practitioner to restate the essential elements of the story. Each story may have one or more themes that need to be restated. The description of the main theme can vary in terms of its level of inference. On the one hand, it might be based on an observation and restatement of the client's story, staying as close as possible to the client's own words and experience of the story. On the other hand, practitioners may wish to move somewhat further away from the person's description of the story into a more *interpretive* or even *diagnostic* level. Elaboration on the story might even be developed by having the person free-associate to elements in the story. However, for persons who are learning the TAT or who use it infrequently, it is recommended that practitioners stay close to the client's own description. It should be as brief as possible and should aim to extract the essence of what has been described.

2. Main Hero/Heroine

The hero/heroine is usually the person who is most frequently referred to in the story. More information is given on his or her feelings, beliefs, and behaviors than on those of any of the other characters. As a result, the client is assumed to be identifying with this person. In some stories, there might be a degree of uncertainty as to exactly who is the hero/heroine. The practitioner should infer that the story character who is most similar to the client in terms of age, sex, and other characteristics is the hero/heroine. In certain rare cases, there may be one or more heroes/heroines. The Analysis Sheet further requests that the clinician rate the hero in terms of *interests, traits, abilities, adequacy,* and *body image*. The adequacy of the hero/heroine refers to an ability to complete tasks in a socially, emotionally, morally, and/or intellectually acceptable manner. This level of adequacy would be directly related to the ego strength of the hero/heroine—or, more inferentially, of the client. The body image refers to the style and qualities with which the body or body representation is depicted. Direct descriptions of the body are usually easy to interpret but a more indirect representation, such as certain symbolical features of the violin in TAT Picture 1, might also be included.

3. Main Needs and Drives of the Hero/Heroine

The *behavioral needs* to be rated in the story refer to the most basic needs expressed in the client's story productions (i.e., affection, aggression, achievement). The descriptions of these needs are fantasy productions by the client and might reflect actual conscious needs as well as more disguised latent needs. The clinician might wish to simply state what the clearest and strongest of these needs are, or make inferences about the actual meaning of these needs for the client. For example, extreme nurturance expressed in the stories might indicate that the client demands nurturance from others, or, conversely, that this is a frequent need that he or she expresses. Another example might be extreme avoidance of aggression, which could suggest that the client has a high level of underlying aggression that is being denied.

Clinicians should also note any *figures, objects,* or *circumstances* that are *introduced* as well as any that have been *omitted* but perhaps should have been included. Particularly noteworthy omissions include these: no mention of the rifle in Picture 8BM, the gun/keys in Picture 3BM, or the seminude in the background of Picture 4, or no sexual references to Picture 13MF. The implications of these inclusions/omissions should also be noted. For example, the inclusion of a relatively large number of weapons, food, and money might suggest high needs for aggression, nurturance, or financial success. The omission of important objects in the story productions might suggest some areas of repression, denial, or anxiety associated with the omitted objects.

4. Conception of the Environment (World)

Clinicians should summarize the most important and strongest conceptions of the person's environment. They might be alerted to this distinction by noting the number and strength of descriptive words such as *hostile, dangerous,* or *nurturing.* The summaries of conceptions of the world might include the overall meaning for the hero/heroine—for example, the environment is overly demanding, a wealth of opportunities, or something to be exploited and used.

5. Figures Seen as . . .

One of the main characteristics of the TAT stories is that they can be seen as "apperceptive distortions of the social relationships and the dynamic factors basic to them" (Bellak, 1993, p. 92). Thus, one of the cornerstones of TAT interpretation is understanding how the client views other persons, as represented in the story productions. This category attempts to elaborate on this by rating the hero/heroine's attitudes and behaviors toward *parental, contemporary* (age-related peers), and *junior figures.* For example, the level of aggressiveness of persons of the same gender might be noted, along with the response(s) of the hero/heroine (assertive, placating, hostile, withdrawing).

6. Significant Conflicts

The major conflicts within the hero/heroine should be noted by reviewing the client's current feelings and behaviors and assessing how congruent these are. In particular, clinicians should note any contrast between the *actual* feelings/behaviors and how the client *should* feel. For example, he or she might be trying to accomplish two incongruous goals such as need for achievement versus need for pleasure, or need for hostility versus need for affiliation. Other important conflicts might be between reality and fantasy, or aggression and compliance.

7. Nature of Anxieties

In addition to significant conflicts, clinicians should rate the nature and strength (✓, ✓✓, ✓✓✓) of the hero/heroine's anxieties in terms of fear of *physical harm and/or punishment, disapproval, lack or loss of love, illness or injury, being deserted, deprived, overpowered and helpless, devoured,* or *other.*

8. Main Defenses against Conflicts and Fears

The clinician is asked to rate the presence and strength of defenses against anxieties and conflicts. This helps to provide a description of the person's character structure. The strength of the defenses can be assessed by noting their frequency both within each story and among the different stories. For example, intellectualization occurring in six of the stories suggests a rigid and excessive defensive style. In contrast, the use of several different types of defenses suggests that the client has a much greater degree of variety and flexibility. One option might be to formally score for denial, projection, and identification using Cramer's (1996) *Defense Mechanisms Manual.*

9. Adequacy of Superego as Manifested by "Punishment" for "Crime"

Clinicians are requested to rate the relative degree of appropriateness, severity, consistency, and extent of delay of any consequences for potentially punishable behavior. Particular note should be made of the relative strength and type of punishment compared to the seriousness of the "crime." For example, a harsh superego would be suggested when minor infractions by story characters result in imprisonment or even death. In contrast, a poorly developed superego would be suggested if few or no consequences occurred for a moderate or severe infraction. A section is also included for noting any relevant behavioral observations of the client, such as stammering or blushing, which could suggest an overly harsh superego.

10. Integration of the Ego

In general, the degree of ego integration is indicated by the quality with which the hero/heroine mediates between different conflicts. This is typically reflected in the effectiveness with which the main character can use interpersonal skills. Specific observations can be made regarding the adequacy, quality, effectiveness, flexibility, and style of problem solving. The overall quality (bizarre, complete, original, etc.) of the thought processes involved should also be rated.

Bellak provides a further unnumbered category for rating the client's intelligence. The traditional classifications of very superior, superior, high average, and so on, are used. An additional section allows an overall rating of the client's level of maturity.

In addition to the more traditional TAT areas described, Bellak and Abrams (1997) have provided scales for rating a client's 12 ego functions (I–XII in Table 11.2). These are based on both the total TAT stories the client has provided and on any relevant behavioral observations. A graph can be created by connecting the ratings summarized on Table 11.2. The 12 functions are briefly defined as follows.

I. Reality Testing

This variable rates the extent to which the client accurately perceives and relates to his or her external environment. It requires an accurate appraisal of both the physical environment and the social norms and expectations, as well as an accurate perception of inner reality testing and level of psychological sophistication. These can be partially

assessed by noting the extent to which the client can articulate needs, feelings, values, and beliefs. Also included would be accuracy in perceiving time and place.

II. Judgment

What is the client's capacity for understanding a situation, particularly where interpersonal relationships are involved, and translating this understanding into an effective, coherent response? An appraisal of social and physical consequences as well as competent forward planning are involved.

III. Sense of Reality of the World and of the Self

Here, the clinician rates disturbances in the client's sense of self, such as dissociative experiences, depersonalization, and *dejà vu.* These also relate to feelings of reality or unreality in the client's perceptions of the environment. In particular, how does this sense of reality/unreality relate to the degree to which the client feels that his or her body parts are well coordinated? Other aspects involved in the rating might be the degree of individuation versus differentiation, the sense of self-esteem, and the extent to which the self is experienced as distinct from others and the external world.

IV. Regulation and Control of Drives, Affects, and Impulses

How direct or indirect is the client's expression of impulses? Can they be appropriately and effectively controlled and delayed? How high a tolerance is there for frustration? Is the client undercontrolled or overcontrolled? Can he or she monitor drives and express them in a modified and adaptive manner? Each of these areas should be considered to make a final rating for this category.

V. Object Relations

To what extent are the client's relationships optimal in that they are appropriately relating to, committed to, and invested in others? What is the typical length of the relationships? What is their overall quality? Any distortions, and the degree to which the client gets his or her needs met, should be noted. How mature is the client and how free from maladaptive interpersonal patterns? To what extent is he or she overinfluenced or underinfluenced by others?

VI. Thought Processes

This category requires a rating of the general adequacy and coherence of the client's thought processes. Thus, careful attention should be given to the level of attention, concentration, memory, verbal ability, and abstract reasoning. Are there any distortions, delusions, or unusual associations? Are there clarity and integration in the thought processes? Is the thinking unrealistic, illogical, and characterized by the intrusion of primary process thinking? For example, obsessive-compulsives might be expected to describe minute details of the cards. In contrast, persons with Alzheimer's disease have been found to use fewer words. They typically describe the pictures rather than tell a story about them, and they frequently lose track of the instructions.

VII. ARISE (Adaptive Regression in the Service of the Ego)

Can the client temporarily lower his or her defenses to increase awareness and help with problem solving? This would allow for a relatively free expression of primary process

thinking in which the client can approach self and others from different perspectives. Another concern is how adequately he or she can later reintegrate and reorganize the insights and new perspectives resulting from the lowered defenses. The task of responding to the TAT can be seen as an opportunity to allow for this temporary regression into fantasy activity, with the goal of helping to reveal, problem-solve, and understand aspects of self. Relevant questions might involve whether the client approaches the task easily or defensively. Are the story productions rich and creative, or constricted and defended? When they do enter into the fantasies, do they become lost and incoherent or are they able to organize the contents effectively?

VIII. Defensive Functioning

This category requires the clinician to rate the extent to which clients' defenses protect them from internal anxiety-provoking impulses and conflicts. Are they excessive, defective, adaptive/maladaptive? Overall, how successful are they? How much anxiety or depression does the person experience? The specific types and strengths of defenses have been summarized in Category 8 of the Analysis Sheet so that clinicians can refer to these previous summaries to obtain useful detail. However, this Category VIII differs in that it is more a global rating of defensive effectiveness, using all sources of information available to the clinician.

IX. Stimulus Barrier

The client's stimulus barrier refers to how reactive a person is to various events (high/low threshold). Is the person hypersensitive to minor criticisms or low levels of stress? Does he or she react to unpleasant situations with responses such as anger, aggression, assertiveness, withdrawal, disorganization, and/or victimization?

X. Autonomous Functioning

To what extent is the client disrupted by certain ideas, feelings, conflicts, or impulses? If the client feels disrupted, how much does this compromise his or her ability to work and socialize independently? Instead of functioning independently, does the client become highly dependent on others to cope, decide, and initiate what to do? In contrast, has he or she been able to develop autonomous behaviors such as adaptive habits, time management skills, or hobbies that help toward functioning relatively independently?

XI. Synthetic-Integrative Functioning

Clinicians must rate the client's ability to actively reconcile difficult needs and conflicts. Are important generalizations and similarities among different ideas, events, and persons perceived? Is there an ability to make necessary compromises between disparate areas of personality and/or interpersonal relationships? How adequately can these integrative abilities be used to work with contradictory behaviors, attitudes, values, and emotions?

XII. Mastery-Competency

This final category requires a rating of the client's overall sense of competency, especially as it relates to the outcome of different story themes. Information useful for this rating might come from a variety of different areas: ability to resolve conflict, quality of ego defenses, ego integrity, creative problem solving, relative degree of rigidity of

defenses, self-efficacy, and degree to which the person has an internal versus external locus of control. One important consideration is whether the client's sense of competency is realistic, given his or her actual abilities and achievements. Some clients might either under- or overestimate their level of competency.

INTERPRETATION

When the scoring has been completed, it should be relatively easy to convert this information into a coherent description of the person. The scoring and interpretation can even be considered the same task. In other words, the practitioner can extend on and elaborate on the scoring to make inferences about the client based on the themes occurring in the narratives. Bellak and Abrams (1997) suggest that the three major levels of interpretation are (a) the descriptive, (b) the interpretive, and (c) the diagnostic. The descriptive level is merely a short repeat of the story, as has been outlined in scoring category 1 of the Analysis Sheet. The interpretive level extends the descriptive level somewhat by an alteration of the descriptive level beginning with "If one . . . [does X, then the outcome will be Y]." For example, a descriptive "interpretation" to Card 1 might be: "Boy is practicing to increase his competence." The interpretive level would be "[If one] practices, then he or she will improve." The diagnostic level is a further extension in that an inference is made about the client. Thus, one might infer that, in the story for Card 1, "The client has a high need for achievement with a high level of self-efficacy."

The core features of the client can be organized in the summary section, which has been previously outlined and is included as part of Bellak's Short Form for scoring and interpretation. It is even possible for a report to be organized around the information noted on the 10 different scoring categories. These might be further integrated into the following three areas:

1. *Unconscious structure and needs:* Derived from categories 1 through 3.
2. *Conception of the world and perceptions of significant persons:* Derived from categories 4 and 5.
3. *Relevant dimensions of personality:* Derived from categories 6 through 10.

Further ratings can be noted for levels of intelligence and maturational level.

These areas of description tend to be fairly abstract and inferential. One technique for balancing these abstract descriptions is to include actual story segments to illustrate the points or principles that are being described. This should effectively provide a more qualitative, concrete, and impactful description of the client

One useful interpretive consideration regarding the TAT stories is that approximately one-third of the stories are likely to be impersonal renditions or clichés of previously heard information. In the protocols of highly constricted, defensive clients, this proportion is likely to be even higher. Because of the impersonal nature of these stories, it is usually difficult to infer the underlying determinants of personality. In contrast, some stories are extremely rich in that they reveal important core aspects of the client.

Yet another consideration is that, even though, for the most part, high, moderate, or low scores on the stories correspond to high, moderate, and low characteristics within

the subject, this is not always the case. For example, Murray found that there was a negative correlation (–.33 to –.74) between *n* Sex on the TAT and *n* Sex expressed in overt behavior. Of final and particular note are the subject's current life situation and emotional state at the time of examination. One of the more important variables that can affect the emotional state of the subject, and, therefore, the test results, is the particular interaction between the subject and the examiner. A sensitive and accurate interpretation can be obtained only if the examiner takes into account the existence and possible influence of all these variables.

RECOMMENDED READING

Bellak, L., & Abrams, D. M. (1997). *The TAT, CAT, and SAT in clinical use* (6th ed.). New York: Grune & Stratton.

Cramer, P. (1996). *Storytelling, narrative, and the Thematic Apperception Test.* New York: Guilford Press.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112,* 140–154.

Chapter 12

SCREENING AND ASSESSING FOR NEUROPSYCHOLOGICAL IMPAIRMENT

An important role in clinical practice is screening and assessing for the presence of possible neuropsychological impairment. This is highlighted by data indicating that 20% to 30% of assessment referrals to professional psychologists relate to information regarding central nervous system (CNS) involvement (Camara et al., 2000). This proportion is likely to be even higher for referrals from psychiatric and neurological settings. Information derived from these sorts of assessments might serve as an early warning sign that, if positive, would then result in a more in-depth medical or neuropsychological assessment and/or further monitoring of the patient. Examples of the types of situations where screening might be important would be among substance-abusing populations, persons exposed to neurotoxic substances, or elderly populations where the distinction between depression and organically based dementia might be crucial (see Stringer & Nadolne, 2000). Additional situations might occur with a school psychologist who is trying to understand why a student is performing poorly, workers' compensation cases in which brain damage might be suspected, or screening for brain damage among psychiatric populations. Each of these situations would require that the assessing clinician be sensitive to the expression of brain impairment, methods of assessing for it, and the patterns of behavioral and test results that would suggest the presence of such impairment.

This chapter provides introductory knowledge and strategies for screening for CNS involvement. It also provides strategies for assessing different domains of neuropsychological functioning. These domains, and the recommended tests used to assess them, are listed in Table 12.1. While assessing these domains according to the recommended procedures still does not entail a full neuropsychological assessment, it provides considerably more depth than that obtained from simple screening instruments. The chapter, therefore, follows the trend in neuropsychology toward greater sophistication of instruments and away from simple screening procedures. It is also based on the belief that professional psychologists will be expected to become increasingly familiar with the well-established field of clinical neuropsychology. The coverage is as comprehensive as possible in the space limitations of a single chapter. Initially, it may seem somewhat technical. However, it is expected that, when clinicians actually work with the instruments and integrate them into a report, they will internalize the information and find it practical. If a more in-depth coverage is required, readers are referred to Groth-Marnat's (2000a) *Neuropsychological Assessment in Clinical Practice: A Guide to Test Interpretation and Integration.*

When appraising clients with suspected CNS involvement, it is important to appreciate that the behavioral manifestation of such involvement is extremely heterogeneous.

Table 12.1 Recommended screening battery for neuropsychological impairment

Function and Test	Estimated Time for Tests Not Covered Elsewhere
Visuoconstructive Abilities	
Bender (Standard or Background Interference Procedure)[a]	5[b]
Block Design, Object Assembly	
Free Drawing Procedures (Draw-A-Person, etc.)	
Mental Activity (Attention and Speed of Information Processing)	
Trail Making[a]	3
Digit Span	3
Arithmetic	
Digit Symbol	
Memory and Learning	
Rey Auditory Verbal Learning Test[a]	12
Bender memory[a]	2
Digit Symbol, Information	
Verbal Functions and Academic Skills	
Controlled Oral Word Association (FAS)	10
Information, Vocabulary, Comprehension	
Similarities (Arithmetic)	
Motor Performance	
Finger Tapping[a]	5
Executive Functions (Behavioral observations, reports by significant others)	
Emotional Status	
Minnesota Multiphasic Personality Inventory (MMPI-2/MMPI-A)	
Beck Depression Inventory (BDI)	
Total estimated additional time:	50

[a] Indicates tests unique to Chapter 12.
[b] Times given only for tests not covered in previous chapters so clinicians will have an estimate of the additional time(s) beyond the standard core battery.

Some brain-damaged persons might have specific signs such as aphasia, neglect of a portion of their visual field, or word-finding difficulties. In contrast, others might have widespread impairments such as a general lowering of cognitive abilities or difficulty regulating their behavior. Deficits might also range in their expression between being extremely subtle to being quite severe. The practical implication is that any one screening test for neuropsychological impairment is likely to assess for a narrow range of abilities. If a client has deficits outside this range, the test is not sensitive to that particular area of difficulties. The result is a high number of false negatives. Indeed, this problem has plagued most screening devices. For example, a test such as the Bender Gestalt is primarily a test of visuoconstructive abilities. Clients with a wide range of other difficulties are

likely to perform quite well on the Bender Gestalt with the resulting danger that the clinician might erroneously conclude they were not organically impaired.

The presence of false negatives (or false positives) depends in part on the "narrowness" versus the "width" of the test. For example, a test that measures a specific function, such as ability to name objects, is quite narrow in its focus. Clients who do poorly on such a test would most likely be experiencing neuropsychological impairment (true positives). However, there are also many persons who, despite being neuropsychologically impaired, do quite well on such a test and may be misclassified as normal (false negatives), because most neuropsychologically impaired persons do not experience object-naming difficulties. The sign of object naming, thus, is too specific. If another test is used that casts a wider net by using more general indicators (i.e., concrete thinking, impaired immediate memory, distractibility), not as many persons with neuropsychological impairment will be missed (few false negatives). However, many people will be labeled brain damaged who are not (many false positives). This is likely to be particularly true for severe psychiatric patients. Indeed, neuropsychological tests have had a notoriously difficult time distinguishing psychotics, especially chronic schizophrenics, from brain-damaged persons because they often appear quite similar on test performance (Mittenberg et al., 1989).

This difficulty does not invalidate the use of psychological tests for neuropsychological assessment. However, it does highlight the importance of being aware of their limitations and being clear on what they do measure. It also suggests that, instead of using one or two tests, a clinician reviewing neuropsychological impairment ideally should use a number of different tests that assess a wide number of domains. These domains might include tests for mental activity (attention and speed of information processing); visuoconstructive abilities; memory and learning; verbal functions, including academic skills; executive functions (observing, directing, and regulating behavior); motor performance; and emotional status. Accordingly, the emphasis in this chapter is to present a series of brief assessment techniques that can be used to screen for, as well as provide greater depth into, the assessment of neuropsychological impairment. In addition, readers are referred to tests in other sections of the book that are relevant for assessing a client's neuropsychological status (primarily portions of the Wechsler intelligence scales). The result is a relatively brief assessment battery that would not be expected to add more than an additional hour to a usual core test battery (WAIS-III/WISC-III, MMPI-2, etc.). The recommended tests also familiarize students and professionals with some of the more frequently used neuropsychological procedures. However, this group of tests still does not assess a sufficiently wide number of areas to be considered a comprehensive neuropsychological battery.

The two general strategies in neuropsychological assessment are a qualitative or pathognomonic sign approach and the use of quantitative cutoff scores. The pathognomonic sign approach assumes the existence of distinctive behaviors indicative of brain damage. Rotations or perseverations are examples of such signs. Additional ones might be aphasias, line tremor, distortions of drawings, difficulty with serial subtraction, clang responses (i.e., ponder meaning "to pound"), neglecting a portion of a visual field (visual neglect), or difficulty distinguishing whether a stimulus is either on the right or left when they are stimulated at the same time (suppressions on bilateral, simultaneous stimulation). In contrast to the sign approach is the use of cutoff scores, which optimally

separates a person's performance into either a brain-damaged or normal range. The use of cutoff scores is a major and even distinctive feature of the Halstead-Reitan Neuropsychological Test Battery (HRNTB; Broshek & Barth, 2000; Reitan & Wolfson, 1993). For example, the Trail Making Test (Part B) requires examinees to connect a sequence of alternating numbers and letters until a section marked "End" is reached. Times greater than 40 seconds for Part A and 91 seconds for Part B indicate performances in the impaired range. Sometimes a combination of both approaches is used. This can be seen on the Bender Gestalt, which requires examinees to draw a series of designs that are presented to them. Clinicians typically note the presence of pathognomonic signs, such as poor closure or line tremor, but also score different aspects of the drawing to develop a quantitative rating. Scores above a certain level indicate impaired performances.

Similar to other psychological tests, moderator variables, such as age, education, premorbid intelligence, and sometimes ethnicity, are related to neuropsychological test performance. For example, using the standard cutoff score of 91 seconds on Trail Making B results in unacceptably high numbers of misclassifications in the impaired range if used with persons over the age of 70. It has thus been recommended that cutoff scores for determining impairment should use norms corrected for age, education, and gender. These are available in Mitrushina, Boone, and D'Elia's (1999) *Handbook of Normative Data for Neuropsychological Assessment* (see also normativedata.com) and Heaton, Grant, and Mathew's (1991) *Comprehensive Norms for an Expanded Halstead-Reitan Battery.* The norms included in this chapter typically take into account age, education, and gender.

HISTORY AND DEVELOPMENT

Neuropsychological assessment as a well-defined discipline began in the 1950s with the work of Halstead, Reitan, and Goldstein in the United States; Rey in France; and Luria in the Soviet Union. In the United States, the experimental and statistical orientation of American psychology was reflected in test design and use. Norms were refined and used for comparison with an individual patient's performance. Optimal cutoff scores were developed to distinguish impaired from normal performances. In particular, the Halstead-Reitan Neuropsychological Test Battery grew out of an original 27 tests that Ward Halstead selected in the belief that they measured cerebral functioning based on "biological intelligence." He reduced these to 10 tests, and Reitan (1955a) later reduced these to 7. Cutoff scores were developed on these tests; and, based on the proportion of tests in the impaired range, an Impairment Index could be calculated.

Early success was achieved with the HRNTB in distinguishing not only the presence of brain damage, but also the location and nature of the lesion (Reitan, 1955a). During the days before sophisticated neuroradiological techniques, this was extremely useful information. These efforts emerged into an emphasis on what has sometimes been referred to as the three Ls of neuropsychology: **L**esion detection, **L**ocalization, and **L**ateralization. In contrast, there was a relative neglect in the study of diffuse impairment in favor of the stronger emphasis on focal involvement.

Concomitant with the developments in the United States was the work of Alexander Luria in the Soviet Union and Rey in France. They relied extensively on close patient

observation and in-depth case histories. They were not so much interested in what score a person might have obtained, but rather why the individual performed in a certain manner. Their work has epitomized the flexible pathognomonic sign or qualitative approach. Rather than developing a series of quantitatively oriented tests with optimal cutoff scores, Luria emphasized a series of "procedures" that he believed would help the client to express relevant behavioral domains. As such, his approach relied far more heavily on clinician expertise and observation than formal psychometric data. Although somewhat controversial (see K. Adams, 1980), these procedures have more recently been formalized and standardized into the Luria-Nebraska Neuropsychological Battery (Golden et al., 1985).

From these early beginnings, two distinct strategies of approaching neuropsychological assessment emerged. One was the comprehensive battery approach epitomized by Halstead and Reitan and formalized into the Halstead-Reitan Neuropsychological Test Battery; the other was a more flexible, qualitative, hypothesis-testing strategy as represented by Goldstein and Luria. Each of these approaches has different strengths and weaknesses (see Bauer, 2000; Jarvis & Barth, 1994; Russell, 2000). The battery approach has the advantages of assessing both strengths and weaknesses for a broad spectrum of behaviors, is easier to use for research, is more extensively normed and researched, can be administered by trained technicians, and is easier for students to learn. Its disadvantages are that it is typically quite time consuming, may overlook the underlying reasons for a client's specific test score, and is more difficult to tailor toward the unique aspects of the client and referral question. The contrasting qualitative hypothesis-testing approach has the advantages that it can be tailored to the specifics of the client and referral question, emphasizes the processes underlying a client's performance rather than a final score, and is quite time efficient. Measurements of a client's strengths, weaknesses, or certain reasons for ambiguous responses can be pursued in more depth according to decisions made by the examiner. Weaknesses frequently attributed to this approach are that, in practice, it focuses on a client's weaknesses, relies too extensively on clinician expertise, is more difficult to research, is not as extensively researched, and provides a narrower slice of a client's domains of functioning.

Despite the preceding somewhat polarized description, two trends indicate an integration of the quantitative psychometric and the qualitative hypothesis-testing strategies. First, in practice, most neuropsychologists use a combination of the strategies. This is supported by surveys of practice, indicating the vast majority of clinical neuropsychological assessment use a "flexible-fixed" battery comprising a relatively short "fixed" or core battery combined with additional flexible tests that could be selected based on the uniqueness of the client and specifics of the referral question (Sweet, Moberg, & Suchy, 2000). The second trend is the development of objective, in-depth, computerized scoring systems that can help clinicians understand the underlying qualitative processes a client makes in responding to test items (i.e., scoring for the California Verbal Learning Test; Delis, Kramer, Kaplan, & Ober, 1987).

Concurrent with the development of the early testing procedures and batteries, there was also an emphasis on brief screening instruments. The Bender Visual Motor Gestalt was one of the earliest of these. It was first developed by Lauretta Bender in 1938 and comprises nine designs that a client is requested to reproduce. A similar but more complex visuoconstructive test was originally devised by Rey in 1941 and expanded by

Osterrith in 1944. It has since become refined and referred to as the Rey-Osterrith Complex Figure Test (Meyers & Meyers, 1996; Visser, 1992). Subjects are first asked to complete the drawing while it is directly in front of them and then requested to make a second reproduction of the drawing from memory. Rey also developed the Rey Auditory-Verbal Learning Test (Rey, 1964), which primarily screens for difficulties with short-term auditory memory. Clients are instructed to recall a series of words that are read to them and then repeat back as many of the words as possible. A final example of an early screening test for attentional difficulties is the Stroop procedure (A. R. Jensen & Rohwer, 1966; Stroop, 1935). This test presents clients with a series of names of colors but written in different color ink from the written name of the color given (see Ponsford, 2000). For example, the name *green* might be written in red ink. The client is then asked to read the list and give the name of the color of the ink (i.e., red) rather than merely reading the word (i.e., green).

A frequent goal of many of the early screening tests was to differentiate between organic and functional difficulties. Thus, a referral question was sometimes expressed in terms of "ruling out organicity" or to "differentiate between organic versus functional causes." More recently, the appropriateness of this goal and the assumptions behind it have been questioned. In particular, there has been a gradual disintegration of the distinction between many functional and organic disorders. For example, early conceptualizations of schizophrenia considered it functional. In contrast, current research supports strong biochemical and structural correlates in a substantial proportion of schizophrenics (I. Wright et al., 2000). A second factor is that advances in neuroradiological and other neurologically oriented techniques have greatly refined the diagnosis of brain damage. As a result, the use of neuropsychological techniques in diagnosis has become de-emphasized. In contrast, referrals from neurologists and psychiatrists are more likely to request information regarding the nature of already identified lesions.

A further change over time has been that, rather than focusing on measurement, there has been greater emphasis on application (Ponsford, 1988; Stringer & Nadolne, 2000). Thus, it is no longer sufficient merely to state that a client is experiencing cognitive deficits in certain areas. Instead, answers to more functionally relevant questions are being required such as the client's employability, responsiveness to rehabilitation, and the need for certain environmental supports (Sbordone & Long, 1996). This can be clarified by considering the differences between impairment and disability. *Impairment* typically reflects normative comparisons and test data. In contrast, the functionally relevant term *disability* more closely takes into account the context of the client including his or her circumstances, environment, and interests. For example, a client might be statistically in the impaired range on tests requiring sequencing, but if his or her occupation required primarily visuospatial skills, he or she might be able to continue functioning effectively. In contrast, a computer programmer who developed an equal level of sequencing difficulties would be likely to become quite disabled by this problem. There are increasing expectations of clinicians to work with both the test data and the specifics of the client to translate the impact of any test-related impairment into a better understanding of the meaning this might have for the client in his or her everyday life. This may also require using methods of analysis other than psychological tests such as the ratings of relatives, ward observation charts, and simulations (Knight & Godfrey, 1996; Sbordone & Guilmette, 1999).

Consistent with these points is that more recent emphasis has been not so much on measuring "organicity" or "brain damage," but rather on assessing different functions or domains. This might include attention, short-term memory, or visuoconstructive abilities. Thus, "brain sensitive" screening tests should not be considered to be tests of brain damage, but rather tests of certain functions that *may* be consistent with CNS involvement. This means that, instead of using single tests, neuropsychological assessment ideally should use several instruments to examine a number of different domains. This emphasis is reflected in this chapter through the presentation and elaboration of a number of short, easily administered tests that cover a wide range of the person's abilities. There have also been a number of formally developed screening and short assessment batteries by other authors. Wysocki and Sweet (1985) developed a seven-test battery composed of Trail Making, finger-tapping speed, drawing a Greek cross, the Pathognomonic Scale of the Luria-Nebraska Neuropsychological Battery, the Stroop, and the Logical Memory and Visual Reproduction subtests of the Wechsler Memory Scale. Total administration time is approximately 60 minutes. Another representative screening system is the BNI Screen for Higher Cerebral Functions (Prigatano, Amin, & Rosenstein, 1992a, 1992b). Its purpose is to determine whether patients are capable of taking other neuropsychological tests; it evaluates their level of self-awareness, provides qualitative information regarding cognitive functioning, and assesses a wide range of cerebral functions. The entire procedure typically takes 10 to 15 minutes to complete. There have also been two abbreviated versions of the Halstead-Reitan Battery by Golden (1976) and Erickson, Caslyn, and Scheupbach (1978).

In addition to these procedures, several short batteries have been developed for reviewing possible neuropsychological impairment with specific types of disorders. Batteries for the evaluation of neurotoxicity are the California Neuropsychological Screening Battery (Bowler, Thakler, & Becker, 1986), Pittsburgh Occupational Exposure Test (C. Ryan, Morrow, Parklinson, & Branet, 1987), and the Individual Neuropsychological Testing for Neurotoxicity Battery (R. Singer, 1990). Similar to the previous screening batteries, each of these uses a combination of previously developed tests such as Trail Making and portions of the Wechsler intelligence scales. Assessment and monitoring of some of the more important domains of dementia might be achieved with the CERAD Battery (Morris et al., 1989) or the Dementia Assessment Battery (Corkin et al., 1986). A similar specialized battery for detecting the early signs of AIDS-related dementia is the NIMH Core Neuropsychological Battery (Butters et al., 1990).

INTERVIEWING FOR BRAIN IMPAIRMENT

While tests can be quite useful, the strongest tool for a clinician can often be a clear, thorough, and well-informed history. One of the major factors guiding such a history is understanding the types of behavior that are likely to reflect neuropsychological impairment (see Sbordone, 2000a). Table 12.2 provides a summary of possible behavior changes indicative of impaired brain processes. While the presence of one of these is not sufficient in and of itself to diagnose pathology, several of them would suggest such a process. An additional tool in extracting the range of possible symptoms is a checklist of potential areas of difficulties that the client can easily complete. This might be

Table 12.2 Behavior changes that are possible indicators of a pathological brain process

Functional Class[1]	Symptoms and Signs	Functional Class[1]	Symptoms and Signs
Speech and language	Dysarthria Dysfluency Marked change in amount of speech output Paraphasias Word-finding problems	Visuospatial abilities	Diminished or distorted ability for manual skills (e.g., mechanical repairs, sewing) Spatial disorientation Impaired spatial judgment Right-left disorientation
Academic skills	Alterations in reading, writing, calculating, and number abilities (e.g., poor reading comprehension, frequent letter or number reversals in writing)	Emotional	Diminished emotional control with temper outbursts, antisocial behavior Diminished empathy or interest in interpersonal relationships without depression
Thinking	Perseveration of speech or action components Simplified or confused mental tracking, reasoning, concept formation		Affective changes without known precipitating factors (e.g., lability, flattening, inappropriateness)
Motor	Lateralized weakness or clumsiness Problems with fine motor coordination Tremors		Personality changes without known precipitating factors Increased irritability without known precipitating factors
Perception	Diplopia or visual field alterations Inattention (usually left-sided; may be perceptual and/or in productions) Somatosensory alterations (particularly lateralized or confined to one limb)	Comportment[2]	Altered appetites and appetitive activities (eating, drinking, play, sex) Altered grooming habits (overly fastidious, careless) Hyper- or hypoactivity Social inappropriateness

[1] Many emotionally disturbed persons complain of memory deficits that typically reflect their self-preoccupations, distractibility, or anxiety rather than a dysfunctional brain. Thus memory complaints in themselves are not good indicators of neuropathology.

[2] These changes are most likely to have neuropsychological relevance in the absence of depression, but they can be mistaken for depression.

Source: From *Neuropsychological Assessment, Third Edition,* by Muriel Lezak, copyright 1995 by Oxford University Press, Inc. Used by permission of Oxford University Press, Inc.

informally developed by a clinician through simply listing all potentially problematic behaviors such as difficulties with memory, hearing, depression, or confusion. Alternatively, a checklist is commercially available to allow clients to detail the full range of their symptoms (Neuropsychological Symptom Checklist; Schinka, 1983). Any items a client endorses can be further explored in the interview to determine when the symptoms began as well as their onset, frequency, intensity, and duration.

A family history should focus on some of the general areas previously outlined in Chapter 3. The family history for neurological and/or psychiatric complaints should receive particular attention. A family history that includes conditions with a known or suspected genetic component such as schizophrenia, early onset Alzheimer's disease, Huntington's chorea, or hypertension should alert the clinician that similar processes may be occurring with the client. The presence of any early deaths in the family, learning disabilities, or mental retardation would also be important to consider. Because some types of clients have difficulty recalling detailed information, relevant family members might be contacted to help obtain, elaborate, or verify information.

Prenatal and early personal history are also important areas for consideration. The client's prenatal environment might have involved relevant events such as his or her mother's exposure to pesticides, solvents, dyes, drugs, or alcohol. Complications during pregnancy and birth, such as low birth weight, forceps birth, premature birth, or difficulties related to any anesthetics used, should also be considered. Early developmental milestones, including the age at which the client sat upright, walked, and talked, should be noted and verified with an outside source. Academic history is particularly helpful in determining the person's premorbid level of functioning. Favorite and worst subjects, grades obtained, and highest level of education are all significant. Assessing for possible attentional or learning difficulties is also essential. School records often provide useful information, especially when objective, and outside support is required to verify a client's claims related to his or her premorbid level of functioning.

A client's occupational history helps establish his or her premorbid level of functioning and social functioning. Each occupation requires certain skills that might have implications in interpreting test results. For example, test scores indicating average verbal skills would mean something quite different for an unskilled laborer than for a successful attorney. Average scores might be consistent with the former but could very well reflect impairment for the latter. It also might be relevant to note whether the person's occupation has resulted in exposure to potentially neurotoxic substances such as organic solvents, insecticides, lead, or mercury. If so, the occupational precautions used and occurrence of all incidents would need to be determined. Knowing current and past interests and hobbies can develop a more complete portrayal of the person.

A review of the client's medical history and any available medical records should be obtained from the client as well as from relevant persons close to the client. The central focus of such a review is to attempt to determine whether the current symptoms can be accounted for based on this history. A person might have had a recent head injury, but inferring that his or her symptoms are partially or wholly the result of this injury might be more difficult. The history might include previous head injuries, high fevers, learning disabilities, or exposure to neurotoxic substances. Any history of a head injury should include details as to the last memory he or she had before the injury, recall of the injury itself, the length of time the person was unconscious, and the

first memory following the injury. Any behavioral changes (irritability, poor memory, confusion) should be carefully documented. Further relevant medical complications might include history of high fevers (103+ F) or significant infectious diseases (meningitis, encephalitis, HIV/AIDS), thyroid dysfunction, diabetes, epilepsy, hypoxia, suicide attempts, hypertension, or neurosurgery for complications such as tumors or aneurysms. If he or she has undergone any surgery, details should be obtained related to anesthetic use, complications, possible loss of consciousness, psychosocial changes following the surgery, and the nature and duration of these changes. Headaches, especially if accompanied by neuropsychological complaints, might suggest a tumor or a vascular disorder. Drug and alcohol use also needs to be carefully documented along with possible changes in prescription or nonprescription medication. Any current or past psychiatric difficulties might also complicate a client's presentation of neuropsychological symptoms.

Any neuropsychological history should provide a careful documentation of present complaints and current overall life situation. Each symptom should be described along with its onset, frequency, duration, intensity, and any changes over time. Asking the client when the symptom first appeared and how it has changed over time frequently can access much of this information. For example, the abrupt onset of neuropsychological complaints with no clear-cut trauma suggests a cerebral vascular accident. In contrast, gradual change might suggest a dementing condition or a slow-growing tumor. Discrete, temporary symptoms suggest transient ischemic attacks. A complicating factor is that clients vary in relation to their awareness of symptoms. Some might be preoccupied with them, others might be indifferent, while still others might be aware of some difficulties but relatively unaware of others. This would then require that the interviewer refer to medical records and relevant persons in the client's life. This would be especially important in conditions such as dementia or frontal lobe impairment in which clients might be both unaware of their deficits and inaccurate regarding details of their personal history (desRosiers, 1992; Gilley et al., 1995). A client's sexual functioning can often reveal relevant information related to neuropsychological status. Changes in sexual desire might be related to certain medications, growth of tumors in strategic areas, affective disorders, infectious diseases, exposure to neurotoxins, or head injuries (especially with frontal lobe involvement). It is also wise for clinicians to investigate the psychosocial factors that might be related to symptoms. Stress, depression, and family turmoil might either cause or serve to exacerbate "neuropsychological" symptoms such as concentration, memory, confusion, and irritability (Burt, Zembar, & Niederehe, 1995; Sherman et al., 2000). Finally, legal complications might be intricately entangled with symptoms. This is especially true for cases involving litigation or workers' compensation.

Whereas the preceding suggestions represent a variety of areas that can be explored flexibly, several structured interviews and questionnaire formats are currently available. The Neuropsychological History Questionnaire (Wolfson, 1985) is an easily completed, 37-page, comprehensive series of questions to be answered by the client. It includes topics such as referral information, academic history, medical and general history, and present status compared with preinjury/preillness status. The Neuropsychological Status Examination (Schinka, 1983) includes a similar organization of topics but is a semistructured interview in which most of the questions

are asked by the interviewer. The Neuropsychological Status Examination also includes the previously mentioned Neuropsychological Symptom Checklist, which provides a brief self-report of symptoms that can be used to assist the interview. An extremely detailed and long (3 to 5 hours) structured questionnaire is the Neurobehavioral Assessment Format (A. Siegel, Schechter, & Diamond, 1996). Additional useful tools might be brief, simple rating forms such as the Mini-Mental State (Folstein et al., 1975), Neurobehavioral Rating Scale (H. Levin et al., 1987) or Patient Competency Rating (Prigatano, 1986). Any of these structured formats requires an examiner to integrate the information into the unique characteristics of the client and relevant test data.

TESTS FOR SCREENING AND ASSESSING FOR NEUROPSYCHOLOGICAL IMPAIRMENT

The criteria for including the tests in this section are that they are brief, well researched, psychometrically sound, frequently used, and collectively provide an overview of a broad range of neuropsychological functioning (see Table 12.1). This selection of tests is similar to recommendations made by other authors for batteries of screening or brief assessment instruments (see Berg, Franzen, & Wedding, 1987; Kane, 1991; Lezak, 1995; Wysocki & Sweet, 1985). A number of the instruments listed in Table 12.1 have already been extensively described in a more general testing context elsewhere in the book (WAIS-III/WISC-III, WMS-III, MMPI/MMPI-2/MMPI-A, BDI/BDI-II). This chapter highlights their use in neuropsychological assessment. In contrast to these more general tests, other tests listed in Table 12.1 are highly specific to neuropsychological assessment and have not been described elsewhere (Bender Gestalt, Trail Making, Rey Auditory Verbal Learning, Controlled Oral Word Association, finger tapping). Accordingly, this chapter provides details about their history and development, psychometric properties, and interpretation.

The various tests in this chapter (see Table 12.1) are organized around domains (visuoconstructive abilities, mental activity, etc.) rather than merely the names of the tests. This increases the likelihood that clinicians will focus on relevant abilities of the client rather than being test oriented. It thus makes it easier to organize relevant topics for inclusion in the psychological report. However, it should be stressed that none of these tests are pure measures of those abilities, which means that clinicians should be cautious in regarding them as such. For example, previous discussions of Digit Symbol (Chapter 5) have stressed that it involves not only attention and speed of information processing but also rote learning, sequencing, and high test-taking motivation. It is then incumbent on the examiner to work with all sources of information to determine the most accurate meanings underlying test scores.

A further important caution and limitation of the chapter is that, while a number of relevant domains have been included, the full assessment of these domains should not be assumed. For example, a number of memory functions have still not been assessed even when all the tests listed in the Memory and Learning section have been given. Thus, a person with memory and learning impairments might still be misclassified as normal. The procedures in this chapter should not be considered a full

neuropsychological evaluation, but the recommended tests can provide far more information than the practitioner can obtain merely using one or two screening instruments.

Although the recommended procedures in this chapter do not entail a full neuropsychological interpretation, the results can be helpful in making tentative predictions about the current level of functioning of the client, degree of adjustment, and likely response to rehabilitation, employment, or interpersonal relationships. These uses are partially supported in that cognitive tests can often be quite accurate in making such predictions (Acker, 1990; Bowman, 1996; Prigatano, 1999; Sbordone & Long, 1996; Vilkki et al., 1994). However, practitioners must be careful to avoid overgeneralization. For example, standard intelligence test scores have not been found to predict psychosocial recovery following closed-head injury with suspected frontal lobe involvement (Vilkki et al., 1994). In contrast, tests of mental flexibility and mental programming (such as the Stroop procedure) were good predictors. Furthermore, measures of emotional functions can often be better predictors of level of disability than cognitive abilities based on neuropsychological test performance. These representative areas of research emphasize the importance of taking into account measures of psychosocial functioning as well as taking a flexible approach to test administration and interpretation.

TESTS OF VISUOCONSTRUCTIVE ABILITIES

The accurate construction of objects involves intact visual perception along with effective visuospatial and visuomotor abilities. Each one of these three areas (perceptual, spatial, motor) might have disturbances that could make visual construction more difficult. Benton (1979) has listed these as follows:

1. *Visuoperceptual disturbances.* Impaired discrimination of complex stimuli, visual recognition, color recognition, figure-ground differentiation, visual integration.
2. *Visuospatial disturbances.* Impaired localization of points in space, topographic orientation, neglect of part of a person's visual field, difficulties with direction and distance.
3. *Visuomotor disturbances.* Defective eye movement, assembling, graphomotor performance.

For some patients, these disturbances might occur together, whereas with others, they might occur separately. A patient might have excellent visuoperceptual abilities but still have significant problems making accurate constructions. At other times, poor perception would lead to or occur in combination with poor constructional abilities. In addition, the ability to draw and assemble objects can be quite variable for a particular patient whose ability to assemble objects might be intact (as in Block Design) but whose drawings of human figures or simpler designs might be quite poor.

Each one of the three disturbances is also likely to have somewhat different neuroanatomical pathways. The practical implication is that any inferences regarding localization of lesion should not be made by merely taking into consideration a person's

overall score on a particular visuoconstructive test. Although overall scores are of limited use, important information and the implications for localization can be derived more appropriately from a careful observation of how the client approaches the task, and the types of errors the person makes. In general, patients with lesions in their right hemispheres tend to approach visuoconstructive tasks in a fragmented, piecemeal fashion in which they often lose the overall gestalt of the design. In contrast, left hemisphere lesion patients are likely to duplicate the overall gestalt of the design but often omit important details of the drawing. Further information on behavioral observations and error types is provided in the sections on interpreting the specific tests of visuoconstructive abilities.

The tests that are recommended and described here should provide clinicians with a good overview of visuoconstructive functions. The Bender Gestalt is a simple, straightforward task that has been extensively researched and frequently used in clinical practice (Camara et al., 2000; Lacks, 1999, 2000). The Bender Gestalt might be further supplemented with a free drawing task such as a Human Figure Drawing or House-Tree-Person. Other somewhat simpler drawing tasks, such as drawing a clock, bicycle, or Greek cross, have also been frequently used. Whereas these tasks involve drawing, the Block Design task requires assembling (rather than drawing) designs with the added factor of a time limit.

Bender Visual Motor Gestalt Test

The Bender Visual Motor Gestalt Test (Bender, 1938), usually referred to as the *Bender Gestalt* or simply the *Bender,* has been extensively used as a screening device for neuropsychological impairment by assessing a client's visuoconstructive abilities. It consists of nine designs that are sequentially presented to subjects with the request that they reproduce them on a blank, 8.5-by-11-inch sheet of paper. The subject's designs are then rated on their relative degree of accuracy and overall integration. Its popularity can be partially accounted for in that it is brief, economical, flexible, nonthreatening, nonverbal, and extensively researched.

Despite sometimes-equivocal reviews and ambiguous research findings, the Bender Gestalt has consistently been one of the five or six most frequently used tests (Camara et al., 2000; Kamphaus et al., 2000). This is consistent with other studies on test usage dating back to 1969. In contrast, Camara et al. found that it was ranked as the 25th most frequently used test by specialty neuropsychologists. This likely results from the greater number of options among specialty neuropsychological tests, along with its not being as highly regarded among this subgroup.

A wide number of scoring systems for adults and children have been developed for the Bender Gestalt, each having various advantages and disadvantages. One of the earliest and most widely accepted scoring systems for adults was developed by Pascal and Suttell (1951). Although this system is widely cited in research studies, it has not gained wide acceptance in clinical settings, primarily because of its complexity and time inefficiency. Another early adult system was developed by Hutt in the 1940s and later formally published in 1960 (Hutt & Briskin, 1960). Although his interest in the Bender Gestalt was primarily for projective personality assessment, he also listed "12 essential discriminators of intracranial damage" (Fragmentation, Closure Difficulty, etc.). Lacks

(1984) adapted this system and provided a detailed scoring manual along with substantial empirical support. In contrast to the Pascal and Suttell (1951) system, it is straightforward and time efficient, typically taking three minutes or less to score. Studies using her system have reported diagnostic accuracies of from 64% to 84% with a mean of 77% (Lacks, 1984, 1999, 2000; Lacks & Newport, 1980). The system is limited to persons 12 years of age or older.

A system for children was developed by Koppitz (1963, 1975). She carried out an extensive standardization of 1,104 children from kindergarten through fourth grade. Her system provides measures of both developmental maturation and neuropsychological impairment. She cautions that, for a diagnosis of brain damage, the examiner needs not only to consider the child's scores, but also to observe the time required to complete the test, the amount of space used, and the child's behavior and relative degree of awareness about his or her errors. The original Koppitz system was developed for relatively young children because the scores of children over the age of 10 no longer correlate with either intelligence test results or age. In addition, after the age of 10 most individuals obtain nearly perfect scores. However, more recent research has indicated that the Koppitz system can be used for adolescents between the ages of 12 and 18 although the relation with age is not nearly as strong as with younger children (McIntosh, Belter, Saylor, Finch, & Edwards, 1988; S. Shapiro & Simpson, 1995). The Koppitz interpretive guidelines are summarized later in this section. A detailed scoring manual can be found in Appendix A of Koppitz's book, *The Bender Gestalt Test for Young Children* (1975).

Whereas the use of the Bender Gestalt in screening for brain dysfunction has been generally accepted, its use in personality assessment has been questionable. Single-sign indicators have rarely been found to be valid. For example, "edging" (consistently drawing the designs along the edge of the paper) has not generally been found to indicate personality variables (Holmes, Dungan, & Medlin, 1984; Holmes & Stephens, 1984). Likewise, projective interpretations that rely heavily on psychoanalytic theory and clinical lore have neither been generally accepted nor sufficiently validated (Sattler, 1985, 1992). However, global ratings that typically sum a series of indicators (size increases, collisions, scribbling, etc.) have had greater validity. For example, accurate discriminations have been made for impulsivity by comparing total scores for impulsive versus nonimpulsive indicators (Oas, 1984). Likewise, Koppitz (1975) has listed emotional indicators that have been found to be good predictors of the general presence of psychopathology when three or more are present (Koppitz, 1975; Rossini & Kaspar, 1987). Thus, the Bender Gestalt has generally been found to be valid in predicting the absence or presence of psychopathology based on clusters of indicators rather than on single signs. With the possible exception of impulsivity and anxiety, the Bender Gestalt is probably ineffective in identifying specific personality characteristics or specific psychiatric diagnoses.

Although the Bender Gestalt has a good track record of achievements, a number of cautions and limitations surround its use. The test has often been described as "assessing" brain damage, yet it is perhaps more accurate to say that it is a "screening" device for brain damage. It does not provide in-depth information about the specific details and varieties of such damage. In fact, the Bender Gestalt is limited to relatively severe forms of brain damage, especially in the right parietal region of the right hemisphere

(Black & Bernard, 1984). Thus, a patient may have significant lesions or subtle deficits that could easily go undetected if a traditional scoring of the Bender Gestalt were the sole method used to assess the presence of cerebral impairment. It is more correct to say, then, that the Bender Gestalt is a screening device for generalized impairment and/or right parietal involvement.

One difficulty in interpreting Bender Gestalt performance is that a certain degree of overlap often exists between emotional and organic indicators on the Bender Gestalt. For example, one of the better indicators for organic impairment is the presence of difficulties with overlapping, which has been found in the Bender Gestalt records of 45% of patients with organic impairment. However, Lacks (1984, 1999) also found that overlapping difficulties occurred in the records of 26% of persons with personality disorders and 26% of those with psychosis. The degree of overlap occurring in the scores of different populations has led some reviewers (Dana, Field, & Bolton, 1983; Sattler, 1985) to seriously question the clinical usefulness of the Bender Gestalt. From a clinical practice perspective, this means that practitioners should not rely on single indicators, such as overlapping or rotations, but rather use optimal cutoff scores to help determine the presence of organicity. If single signs are attended to, they should be used tentatively to formulate hypotheses or to provide qualitative descriptions of specific difficulties in performance.

A further difficulty with the Bender Gestalt is the absence of a commonly accepted and verified scoring and interpretation system. The result is that different research studies have often used different systems, which makes it somewhat difficult to compare their conclusions. Clinicians generally begin by learning a system of scoring and interpretation, but end up with their own unique, subjective approach based on clinical impressions. Although this may be a highly workable, flexible approach, disagreements between "experts" can occur because of their differences in approaching the designs. Another difficulty in depending on clinical impressions is continued, unwarranted reliance on unsubstantiated and possibly incorrect clinical "lore." Lacks (1984, 1999) has argued for the use of formal scoring by presenting evidence that clinicians could increase their diagnostic accuracy for organic impairment on the average of 10% to 15% by using a brief, easily learned, objective scoring system.

Reliability and Validity

Reliabilities across the Lacks adaptation and the Koppitz system have been generally good. Using the Lacks adaptation of the Hutt system, interscorer reliability for the 12 organic signs has been found to range between .87 and .90 (Lacks, 2000). Percentage of agreement on scores ranged from 77% to 86%, and the agreement for diagnosis (brain versus non-brain-damaged) ranged between 84% and 94% (Lacks, 1999, 2000). Test-retest reliabilities over a 3- to 12-month interval were .79 for protocols from neuropsychiatric patients, .66 for patients with Alzheimer's disease, and from .57 to .63 for older adults (Lacks, 1984, 1999, 2000).

Interscorer reliabilities for the Koppitz system have been found to be excellent (.88 and .96), although test-retest reliabilities over a four-month interval were somewhat low (.58 to .66; Koppitz, 1975; Neale & McKay, 1985). Overall, the test-retest reliabilities for the Koppitz system range from .53 to .90 ($Mdn\ r = .77$), depending on age and time between retesting (Koppitz, 1975; Neale & McKay, 1985). The test-retest reliability for

the total number of errors was .83, but reliabilities for specific errors (distortion, rotation, integration, perseveration) were too low to be dependable. Thus, the major focus should be on the total-error score rather than the specific features of the reproductions.

In many studies, the Bender Gestalt has been able to demonstrate its ability to discriminate brain-damaged from non-brain-damaged populations (Hain, 1964; Lacks, 1984, 1999, 2000; Marley, 1982). Studies using the Lacks adaptation have reported diagnostic accuracies of from 64% to 84% with a mean of 80% (Lacks, 1999, 2000; Lacks & Newport, 1980). Its diagnostic accuracy has been questioned, however, when used to assess subtle neuropsychological deficits, such as among many epileptics, or when a differentiation is attempted between functional psychotic patients and brain-damaged patients (Hellkamp & Hogan, 1985). The differentiation between brain-damaged and psychiatric patients has been found to be particularly difficult when distinguishing severely disturbed chronic schizophrenics from brain-damaged patients. However, this distinction may be inappropriate, because schizophrenia is being progressively more conceptualized as an organically based disorder. Further studies have found that Bender Gestalt performance has been able to differentiate Alzheimer's patients from controls as well as reflect the progression of the disease (Storandt, Botwinick, & Danzinger, 1986). Similarly, Bender Gestalt scores were able to predict the extent to which head trauma patients could function independently (Acker & Davis, 1989).

The hit rate with the Lacks adaptation compares favorably with other Bender Gestalt scoring systems. Under equal conditions, the Lacks adaptation had an 84% hit rate, whereas the Pauker was 79% and Hain was 71% and, when scoring only for rotations, the hit rate was 63% (Lacks & Newport, 1980). The Lacks adaptation also has a lower rate of false negatives than either the Hain or Pauker systems. In addition, the Lacks adaptation compares favorably with the Halstead-Reitan composite impairment index, which has a hit rate of 84% for identifying organic impairment (Dean, 1982; Reitan, 1974a). Despite these similarities in diagnostic hit rates, the Bender Gestalt takes only three to eight minutes to administer, whereas the tests specific to the Halstead-Reitan can take up to three to four hours. For screening purposes, the Bender Gestalt has a clear advantage simply because of its greater time efficiency. However, the Halstead-Reitan, when used in combination with the WAIS-III, provides detailed information on a wide range of cognitive strengths and weaknesses.

The validity of Koppitz's (1975) developmental system depends primarily on the purpose for which it is used. Validity is rather good as an index of perceptual-motor development because error scores decrease with age, between the ages of 5 and 9 (Koppitz, 1963, 1975). Concurrent measures of visual-motor perception also suggest a moderate level of validity based on correlations with the Developmental Test of Visual Motor Integration (*Mdn* $r = .65$) and the Frostig Developmental Test of Visual Perception (*Mdn* $r = .47$; Breen, 1982; D. Wright & DeMers, 1982). Correlations with intelligence and academic achievement have been low to moderate (Koppitz, 1975; Lesiak, 1984; Vance, Fuller, & Lester, 1986). For example, correlations with the WISC-R performance subtests ranged from .51 (Block Design) to .08 (Coding; Redfering & Collings, 1982), which suggests that the quality of Bender Gestalt performance is moderately related to ability to perform well on Block Design but not on Coding. Moderate correlations (.57) have also been found between the K-ABC Simultaneous Scale and Bender Gestalt error scores (Haddad, 1986). Significant correlations have also been reported between first

graders' Bender Gestalt scores and their level of performance in reading and arithmetic (P. Ackerman, Peters, & Dykman, 1971; Koppitz, 1958). However, these correlations are sufficiently low so that the Bender Gestalt should not be used as a substitute for a formal intelligence test or a standardized test of academic achievement.

Administration

When administering the Bender Gestalt, the examiner presents the cards directly in front of the client one at a time. Clients are asked to copy each design with a number 2 pencil on a single, blank, 8.5-by-11-inch sheet of white paper that has been presented to the client in a vertical position. A sharpened backup pencil should be available in case a client breaks the graphite on the pencil. Pencils should include erasers. The following verbal directions are taken from Hutt (1985) and are recommended as a standard procedure:

> **I am going to show you these cards, one at a time. Each card has a simple drawing on it. I would like you to copy the drawing on the paper, as well as you can. Work in any way that is best for you. This is not a test of artistic ability, but try to copy the drawings as accurately as you possibly can. Work as fast or as slowly as you wish.** (p. 64)

After the person has completed it, the next one should be presented until the entire nine designs have been reproduced. No comments or additional instructions are to be given while clients are completing the drawings. If clients ask specific questions, they should be given a noncommittal answer, such as, "Make it look as much like the picture on the card as you can." If clients begin to count the dots on Design 5, the examiner may say, "You don't have to count the dots, just make it look like the picture." If they persist, this may show perfectionistic or compulsive tendencies, and the behavioral observation should be considered when evaluating the test results and formulating diagnostic impressions. Although examinees are allowed to pick up the cards, they are not allowed to turn them unless they are in the process of completing their drawing. If it looks as if they have turned the design and are beginning to copy it in the new position, the examiner should straighten the card and state that it should be copied from this angle. As many sheets of paper may be used as desired, although clients are presented with only one sheet initially. There is no time limit, but it is important to note the length of time required to complete the test, as this information may be diagnostically significant.

When clients have completed their drawings, they should be asked to write their names and the date on the paper. No instructions are included as to where this information should be placed and, if asked, it should be clarified that it is up to him or her (Lezak, 1995). A somewhat simpler variation of the preceding instructions may be given to children.

In addition to the standard procedure for using the Bender Gestalt, a memory task is often diagnostically useful. Clients, immediately after having first copied the designs during the standard administration, are asked to reproduce as many designs as possible from memory. This variation provides an assessment of their level of short-term, visual-motor recall. Typically, adult brain-injured subjects are not able to recall the designs as

well as persons who are non-brain-injured (Lyle & Gottesman, 1976). Tolor (1956) found that organic patients could recall an average of only 3.69 designs, whereas convulsive (epileptic) patients and patients with psychological difficulties successfully recalled an average of 5.5 and 5.53 of the designs accurately. These norms were quite similar to those later reported by other authors (Hutt, 1985; Schraa, Jones, & Dirks, 1983). The recall method has not been found to be successful for the screening of children (Kopptiz, 1975). A further variation for memory assessment is to present clients with each design for five seconds, remove it, and then have them reproduce it from memory. A difficulty with the memory procedures is that administration has not been standardized, and scoring criteria have not been developed for what should be considered an accurate level of recall. Thus, clinicians need to rely on clinical judgment, frequently resorting to methods such as giving half credit for partially recalled designs.

An important addition to Bender Gestalt administration procedures for adolescents and adults is the Background Interference Procedure (BIP; Canter, 1983; Heaton, Beade, & Johnson, 1978). This requires the subject to first complete a standard administration and then complete the Bender Gestalt designs on a specially designed sheet of paper that contains a confusing array of curved, intersecting lines. Subjects are not allowed to turn either the card or the paper. Scoring is based on the Pascall and Suttell system. It has been demonstrated that brain-damaged patients show significant decrements in their BIP performance compared with their performances using a standard administration (Canter, 1983; Norton, 1978). This is in contrast to functionally disordered patients and normals who typically do not show significant differences between the two administration procedures. A review of 94 studies by Heaton et al. (1978) indicated that, overall, it had a median 84% correct classification rate in differentiating organic and psychiatric patients.

Interpretation Guidelines: Adults

Quantitative scoring of organic indicators can be obtained by using the Detailed Scoring Instructions in Chapter 6 of Lacks (1999, pp. 67–96). The results can be summarized and tabulated on the Bender Gestalt Test Scoring Summary included in Table 12.3 and also in Lacks (1999, p. 70).

Sometimes the style and manner of drawing, including behavioral observations, can help to develop hypotheses related to personality characteristics. Such observations can assist in providing a context in which to understand the examinee's approach to the task. This underscores the issue that different clients might have the same score/error but do so for different reasons. For example, a rotation may result from a neurologically based processing deficit in one person, but for another it may result from a functionally based sense of disorientation. Another example may occur when a person with a hesitant, obsessive style but with no indication of brain damage takes considerable time (greater than five minutes) to complete the drawings. Another person with documented brain damage who also takes a greater than average time but insists on counting each dot precisely may be attempting to compensate for his or her impairment by developing obsessive behaviors. Other factors that might influence Bender Gestalt performance are situations that might encourage faking, chronic schizophrenia, older age, or a history of substance abuse (Lacks, 1999, 2000).

Table 12.3 Scoring checklist for the Lacks indicators of organicity

_____	1. Rotation
_____	2. Overlapping difficulty
_____	3. Simplification
_____	4. Fragmentation
_____	5. Retrogression
_____	6. Perseveration
_____	7. Collision or collision tendency
_____	8. Impotence
_____	9. Closure difficulty
_____	10. Motor incoordination
_____	11. Angulation difficulty
_____	12. Cohesion
_____	Time greater than 15 minutes

Total score _____

Test diagnosis _____

Source: Adapted from *Bender Gestalt Screening for Brain Dysfunction,* 2nd ed., 1999, by Lacks, New York: Wiley.

The more serious indicators outlined by the different scoring systems (impotence, motor incoordination, angulation difficulty) are most characteristic of brain-damaged populations. However, these can also occur in the protocols of emotionally disturbed persons. Distinguishing between the two categories of disorders based on Bender Gestalt responses can often be difficult. This is further complicated because organically impaired persons usually have emotional responses to their deficits. It is often difficult, if not impossible, to differentiate precisely the extent to which their current problems are organic as opposed to functional. Related to this is that schizophrenics may have a far greater number of brain-related changes than was previously believed (Wright et al., 2000), which again makes a precise division into organic versus nonorganic categories difficult and sometimes inappropriate.

The quantitative method of scoring adult Bender Gestalts using the Lacks adaptation of the Hutt-Briskin scoring system is a relatively brief and straightforward procedure. The 12 "essential discriminators of intracranial damage" outlined in the Lacks adaptation were originally derived from Hutt and Briskin (1960). Descriptions and examples of each of these categories are included in Lacks's (1999) Detailed Scoring Instructions. A computer program is also available to assist with scoring and interpretation of protocols (Lacks, 1996). A brief description of the categories follows. When relevant research is available, some of the descriptions are also accompanied by a brief discussion.

1. ***Rotation: Severe.*** Using the Lacks (1999) system, rotations are scored when there is a change in the orientation of the axis of the figure ranging from 80 to 180 degrees. Mirror imaging of the figure is included as part of the scoring for rotations.

Other scoring systems have been more lenient in defining a rotation as a reorientation of 45 or more degrees, but Lacks scores rotations of 45 to 80 degrees as "Angulation Difficulty" (see Lack's description of Angulation Difficulty).

Research on past systems has found that rotations occur most frequently on Designs 3 (28% of the instances) and A (17%) and least frequently on Designs 6 (2%), 2 (5%), and 1 (6%; Freed, 1969). Rotations occur both for organic and nonorganic psychiatric patients, such as hospitalized persons with intellectual disabilities (Silverstein & Mohan, 1962) and for schizophrenics (Hutt, 1985; Mermelstein, 1983). Thus, differentiated diagnosis between the two groups cannot be made based on the presence of rotations, or any other single sign. Lacks (1984, 1999) has reported that 26% of persons with organic dysfunction made rotations, whereas only 13% of psychotics and 9% of personality-disordered patients did. It has been noted that organics produce more spontaneous rotations, but they also have more difficulty in creating a rotation when specifically requested to do so (Royer & Holland, 1975). This suggests that assessment of the relative difficulty a person experiences in making deliberate rotations may have relevance for differential diagnosis. The primary mental functions associated with rotations are impaired attention, limited capacity for new learning (Marley, 1982), and disorientation (Mermelstein, 1983). As with other visuographic disabilities, the most likely area of the brain to be affected is the parietal lobe (Black & Bernard, 1984). Although Bender Gestalt rotations can occur with either right or left hemisphere lesions, the incidence is about twice as frequent for right hemisphere patients as for left (Diller et al., 1976).

2. ***Overlapping Difficulty.*** If a client has difficulty drawing portions of figures that should overlap, overlapping difficulty should be scored. This might include failing to draw portions that are supposed to overlap, simplifying the figure in the area they overlap, or sketching or redrawing overlapping portions.

3. ***Simplification.*** Simplification is scored when the figure is drawn in a simplified or easier form. Note that simplification is scored only if the figure is drawn in a simpler but not a maturationally more primitive form. If a more maturationally primitive figure is drawn, it is scored for Retrogression (see description of Retrogression). Examples of simplification include drawing very simplified figures, making circles for dots (Design 1), drawing parts that should overlap as being separate (Designs 6 and 7), and making parts that should join as being one-eighth inch or more apart (Designs A, 4, 5, or 8).

4. ***Fragmentation.*** Scoring for fragmentation should occur if the figure is broken up into different parts sufficiently to destroy the actual gestalt of the design.

5. ***Retrogression.*** Scoring for retrogression is made when the figure is drawn in a more primitive gestalt than the actual stimulus figure. Examples include making loops for circles (Design 2), dashes for dots (Design 1), or making a square instead of a diamond (Designs A and 8).

6. ***Perseveration.*** In the more general neurological and psychiatric literature, *perseveration* refers to the continuation of a response well beyond the required number expected. On the Bender Gestalt, Lacks (1999, 2000) describes two types of perseverative errors. Type A, or interdesign perseveration, occurs when there is the "inappropriate substitution of the features of a preceding stimulus, such as replacing the circles of Figure 2 with the dots of Design 1 . . ." (p. 96). Type B, or intradesign perseverations, occurs when the client continues to draw a figure beyond the limits specified by the stimulus. This might occur if 14 or more dots are drawn for Design 1 or 13 or more

columns of circles for Design 2. The presence of perseverations is scored if either Type A or Type B perseverations are present.

Because the general behavior of perseveration has been extensively discussed as being an important feature of some patients who are organically or psychiatrically impaired, it has similarly been researched in relation to Bender Gestalt responses. Although perseveration has been found to exist both in schizophrenic and organic populations (see Tolor & Brannigan, 1980), it is more strongly associated with organicity (Hain, 1964; Lacks, 1999; Lerner, 1972). Lacks (1984) reported that perseveration occurred in 56% of the protocols of brain-damaged patients, whereas only 31% of personality-disordered persons and 32% of psychotics had perseverations on their protocols. Thus, it is not a particularly effective discriminator between organics and other patient groups.

The presence of perseveration suggests impaired executive abilities in which the person may have deficits in initiating, inhibiting, sequencing, and monitoring his or her behavior. To confirm this, check with family members to determine the extent to which the client requires prompting, supervision, and coaching. Behavioral observations during the testing can also be extremely helpful.

Marley (1982) has expanded the definition of perseveration and divided it into three different types. While each type is considered an indicator of organic impairment, she also associated different areas of mental functioning with each one. Marley's Type A perseveration occurs when numbers, letters, or other shapes are substituted for those elements found in the original Bender Gestalt design. This suggests a loosening of associations, impaired planning, diminished attention, poor concentration, and a difficulty with immediate and delayed memory. It is characteristic of dementia and is associated with frontal, frontotemporal, or bilateral involvement. Type B perseveration occurs when additional elements are drawn into Designs 1, 2, 3, and 5, or when additional curves are included in Design 6 (the same as the Lacks adaptation's definition of perseveration). Possible areas of mental functioning are an inability to shift-set, dissociation from the task, diminished attention, poor concentration, concrete thinking, and perseverating behavior outside the testing situation. This is characteristic of dominant hemisphere temporal involvement. The final form of scorable perseveration (Type C) occurs when the examinee redraws his or her design without any effort to erase or cross out the previous one. This can be the result of impaired concentration, intermittent confused ideation, difficulty with planning, and impaired visual-motor functions. Type C perseveration is characteristic of cortical impairment in the parieto-occipital areas of the dominant hemisphere. These findings represent possibilities for determining the nature of cognitive impairments and the location of lesions, which may either support other data or point out future directions to explore.

7. ***Collision or Collision Tendency.*** Scoring for collision is given when the figure is drawn so that it overlaps or collides with another figure.

8. ***Impotence.*** Impotence occurs when clients realize they have drawn the figure incorrectly but appear unable to correct the error. This may be reflected in either their verbal acknowledgment of difficulty or through repeated unsuccessful attempts to draw it accurately. Lacks (1999) found that impotence was a good discriminator between organics and other groups. Whereas 24% of her sample of organics experienced impotence, it was found in only 2% of the records of personality disorders and 4% of psychotics.

9. ***Closure Difficulty: Marked and Persistent.*** This category is scored when the client "shows continuing difficulty in getting parts of figures to join that should join" (Lacks, 2000, p. 419); for example, closing circles and hexagrams or difficulty joining the two portions of Design A. This was the most frequent error among Lacks's (1999) population of organics, with 79% of them demonstrating closure difficulties. The percentages were also somewhat high for psychotics (53%) and personality disorders (55%) indicating that, although this scoring category occurred frequently among organics, it was still not an effective discriminator in and of itself.

10. ***Motor Incoordination.*** Lacks (2000) defines this category as the figures being "drawn with irregular rather than smooth lines indicating tremulousness" (p. 420). This category has been found to rather effectively discriminate between Lacks's (1999) samples of subjects with personality disorders, psychosis, and organic dysfunction. Fully 55% of organics had motor incoordination, whereas it occurred in only 24% of personality-disordered subjects and 13% of psychotics.

11. ***Angulation Difficulty.*** Angulation difficulty is scored when there is "difficulty in producing the angles of Design 2 and 3" (Lacks, 2000, p. 420). For example, Design 2 might be tilted or rotated 45 to 80 degrees (more than 80 degrees would be scored as a Rotation). Lacks (1999) found that angulation difficulty occurred in 41% of her sample of organics, but only in the records of 15% of personality disorders and 18% of psychotics.

12. ***Cohesion.*** Cohesion is scored when there are "isolated decreases or increases in the size of a figure in relation to the other figures in the protocol" (Lacks, 2000, p. 420). For example, cohesion is present when the right portion of Figure A is decreased by more than one third of the size of the portion on the left side.

After a protocol has been scored using Lacks's (1999) Detailed Scoring Instructions, a clinician can then check to see if the examinee's score falls within the brain-damaged range. Lacks (1996, 1999, 2000) gives the normal range as 0 to 4 and the optimum cutoff for organic impairment as 5 or more errors (see Table 12.4). The clinical utility of the Lacks adaptation for the Bender Gestalt can be evaluated by its ability to differentiate organic populations from normals and psychiatric populations other than organics. The hit rate must exceed the typical base rate of 20% to 30% organics and 20% to 30% schizophrenics found in most psychiatric settings. Lacks (1999) presents evidence that it is unusual for nonorganic persons to have 5 or more essential discriminators (error categories). Table 12.4 indicates that 74% to 96% of nonorganics scored less than 5, whereas only 18% of diagnosed organics did. Using a cutoff score of 5 or more indicators results in a hit rate for accurately identifying organics that ranges from 82% to 86% (Lacks, 1999; Lacks & Newport, 1980). McIntosh et al. (1988) suggest that the Lacks adaptation can also be used for adolescent populations from ages 12 to 16 by using a similar cutoff score.

Whereas some of the scoring categories described earlier were relatively more effective at discriminating between organics and other patient groups, none of them should be used as single signs in and of themselves. The only clearly validated approach is to use cutoff scores. However, the presence of some of the different scoring categories, along with relevant behavioral observations, can be used to form tentative hypotheses

Table 12.4 Percentile distributions of BGT total scores for various comparison groups

Number of Errors	Nonpatient Adults ($N = 495$)	Nonpatient Older Adults ($N = 334$)	Nonimpaired Psychiatric Inpatients ($N = 264$)	Psychiatric Inpatients with Brain Damage ($N = 85$)
0	20	5	3	0
1	51	17	10	0
2	75	35	31	4
3	87	52	55	9
4	96	74	74	18
5	98	85	85	51
6	99	92	93	71
7	100	96	96	80
8		99	99	87
9		99	100	95
10		100		99
11				100
12				

Note: The cutoff score for dysfunction is five or more errors.

Source: From *Bender Gestalt Screening for Brain Dysfunction,* by P. Lacks, p. 107. Copyright © 1999 by John Wiley & Sons. Reprinted by permission of John Wiley & Sons.

concerning client functioning. In particular, there are often qualitative differences in the performance of persons with lesions in different areas of the brain. Whereas right hemisphere patients are more likely to make errors related to visuospatial abilities (e.g., rotations, asymmetry, fragmentation, unrecognizable drawings, unjoined lines), persons with left hemisphere lesions often make drawings that are shaky (line tremors) and smaller in size, with rounded corners and missing parts (oversimplification; Filskov, 1978). However, the Bender Gestalt is still more likely to miss patients who have left hemisphere lesions.

Interpretation Guidelines: Children

The Developmental Bender Gestalt Test scoring system (Koppitz, 1963, 1975) is the dominant system used for children. The primary focus is on understanding children's visual-motor abilities as they relate to developmental maturation. Koppitz (1963, 1975) also lists typical errors associated with emotional indicators and brain damage, but places these in the context of what would be expected for a particular individual having a specified chronological age. Bender Gestalt protocols are scored based on the presence of 30 mutually exclusive items. Composite scores can thus range from 0 to 30. The system is relatively easy to learn. The following section is primarily a summary of Koppitz's approach and provides general interpretative guidelines based on indicators for developmental maturation, organicity, visual-motor perception difficulties, and emotional indicators.

The specific scoring criteria developed by Koppitz (1963, 1975) for developmental level can be found in Appendix B of *The Bender Gestalt Test for Young Children* (1975). Table 12.5 in the present chapter summarizes the scoring criteria. A different set of scoring criteria for emotional indicators has also been developed by Koppitz and appears in *The Bender Gestalt Test for Young Children* (1963). To obtain specific scores, clinicians should consult these criteria and use the outline in Table 12.5 as a scoring guide. Both texts by Koppitz (1963, 1975) include important guidelines, cautions, and reviews of research; and clinicians are encouraged to consult these for further elaboration and discussion.

Table 12.5 Summary and scoring sheet for Koppitz developmental scoring system

Design		Type of Error	Check if Present
A	1a	Distortion of shape	_____
	1b	Disproportion	_____
	2	Rotation	_____
	3	Integration	_____
1	4	Circles for dots	_____
	5	Rotation	_____
	6	Perseveration	_____
2	7	Rotation	_____
	8	Row added, omitted	_____
	9	Perseveration	_____
3	10	Circles for dots	_____
	11	Rotation	_____
	12a	Shape test	_____
	12b	Lines for dots	_____
4	13	Rotation	_____
	14	Integration	_____
5	15	Circles for dots	_____
	16	Rotation	_____
	17a	Shape test	_____
	17b	Lines for dots	_____
6	18a	Angles in curves	_____
	18b	Straight line	_____
	19	Integration	_____
	20	Perseveration	_____
7	21a	Disproportion	_____
	21b	Incorrect angles	_____
	22	Rotation	_____
	23	Integration	_____
8	24	Incorrect angles	_____
	25	Rotation	_____

The Bender Gestalt can provide information about a child's perceptual maturity, degree of emotional adjustment, or extent of neuropsychological impairment. A particular clinician may wish to assess only one of these areas, or may consider them all. However, if all the areas are considered, it is crucial to be cautious of the possible overlap among them because many of the signs occur throughout the different scoring guidelines.

A suggested sequence in approaching the Bender Gestalt is to initially develop a global impression of the relative quality of the reproductions as a whole. It is also important to note any relevant behavioral observations made while the child is completing the designs. These observations might include areas such as the child's level of confidence, awareness of errors, completion time, and any comments that are made. The clinician might then look at specific features of the drawings, including figure size, placement, line quality, order and organization of the designs, distortions, erasures, reworking, omissions, and any other unusual treatment. Finally, objective scoring can be made for developmental maturity, organicity, visual-motor perception, and emotional difficulties.

Developmental Maturation As is true with all areas of development, visual-motor perception skills increase with the growth of the child. Although children mature at different rates, the following guidelines developed by Bender (1938) and outlined by Clawson (1962) describe the typical pattern of visual-motor development.

Typical Patterns of Reproduction by Age

Age

2 Has not developed the skills necessary to reproduce the designs with any degree of accuracy but is able to keep pencil on paper and make scribbles, dots, and dashes.

3 Ability to draw loops, lines, arcs, and circles.

4 Can arrange circles or loops in a horizontal left-to-right direction.

5 Figures characterized by having a square appearance; can create many different designs; crosses horizontal and vertical lines.

6 Can create a relatively accurate reproduction of the Bender Gestalt designs, because visual perception is more mature and can be integrated with kinesthetic and tactual perception. Designs A, 1, 4, and 5 are likely to be particularly accurate.

7 Good ordering of designs, relatively accurate reproduction of oblique lines. Subparts to Designs A and 8 are joined. There are no major additions to the child's drawing ability beyond the age of 7, but there is an increasing number of successful reproductions, greater combinations of basic forms, and more refinement in techniques.

8 Accuracy in joining subparts and making dots and an improvement in the contours on curved figures. Design 2 is drawn with vertical rather than oblique columns; Design 3 has columns of arcs instead of angles; figure is accurate except for an obliqueness in vertical support.

9 Less frequent occurrence of rotations; subtle improvement in the detail of the designs, no longer tends to draw the designs vertically.

10 Accurate hexagons drawn for Design 7; subparts correctly joined; oblique columns drawn for Design 2.

11 All designs reproduced accurately with correct sequence, organization, and size.

Visual examples of the preceding maturational guidelines are provided in Appendix J on page 697. The table in Appendix J lists the relevant ages in the left-hand column and the specific designs on the top row. The percentage in each box represents the percentage of persons from a particular age group who produce an accurate reproduction of the designated design. As can be seen, the percentage of accurate reproductions gradually increases for Design A until the age of 11, at which time 95% of all children produce an accurate design. A clinician can develop a rough indication of the person's maturational level by referring to the visual "norms" included in Appendix J.

A more specific rating of developmental level can be determined for children from 5 to 12 by scoring with the criteria developed by Koppitz (1975) and outlined in Table 12.5. However, the decrease in errors with age is not even and steady but rather decreases rapidly around age 8. This results in a skewed distribution, with most of the errors occurring between the ages of 5 to 8 (Taylor, Kaufman, & Partenio, 1984). Thus, it should not be considered developmental past the age of 8 or 10. If there is a significant lag between the child's chronological age and the level at which he or she reproduces the Bender Gestalt designs, the possible causes should be explored with a more complete evaluation of the protocol as well as a review of other relevant data.

Indicators for Organicity When screening for neuropsychological impairment using the Bender Gestalt, it is important to be aware that many of the indicators for central nervous system (CNS) involvement are also indicators for emotional disturbance. This raises the serious possibility of misclassification. Thus, the results of the Bender Gestalt alone are rarely sufficient to make a differential diagnosis between neuropsychological impairment and emotional disturbance; additional information is needed to determine both the nature and cause of the individual's problems.

The error categories described in this section have been reported in the literature to be significant indicators of neuropsychological impairment both for children and for adults. Although the presence of these errors may indicate impairment, even if none of these factors are present, the person may still be suffering a neuropsychological impairment. Conversely, a poor Bender Gestalt performance may reflect a variety of factors, only one of which is neuropsychological impairment. The most common errors associated with organicity are fragmentation or omission of parts of a design, closure problems, distortion of figures, rotations of all or part of a design, perseveration in one design or from one design to another, and substitution of lines for dots. The presence of only one of these is rarely likely to indicate neuropsychological impairment. However, the likelihood of organic deficit increases with the presence of several indicators.

The assessment of neuropsychological impairment using the Bender Gestalt requires a number of important considerations and cautions. Perhaps the most clear and specific indicator is a score greater than one standard deviation above the mean normative

score for a given age group (see Table 12.6). However, alternate considerations should still be made. The high score might be the result of emotional factors, poor motivation, fatigue, a poor understanding of the instructions, or developmental delay. A high score that is caused by developmental delay may merely reflect individual differences in the rate of maturation. The possibility then exists that significant improvement might occur with increased age. Koppitz (1975) points out that there is little to be gained by scoring for both developmental level and brain damage, because they are both equally effective in detecting neuropsychological impairment. This overlap between the two scoring guidelines also underlies the difficulty in differentiating between developmental delay and a more pathological injury to the brain.

When using the indicators of brain injury described and listed in Table 12.7, clinicians should note the overall number of indicators. An abbreviated list of nine indicators follows and can be used as a brief reference and summary. Whereas normal children from the ages of 5 to 8 might be expected to make some of these errors, persons above the age of 8 or 9 should be expected to have few or no errors (Taylor et al., 1984). If four or more of the following characteristics are present, CNS impairment is a strong possibility:

1. Simplification of two or more figures to a level three or more years below the child's chronological age.
2. Collision of a figure with another figure or a reproduction in which a figure runs off the edge of the paper.
3. Fragmentation of one or more figures.

Table 12.6 Distribution of Bender test mean scores and standard deviations

	1964 Normative Sample[1]			1974 Normative Sample[2]		
Age Group	*N*	Mean	*SD*	*N*	Mean	*SD*
5–0 to 5–5	81	13.2	3.8	47	13.1	3.3
5–6 to 5–11	128	10.2	3.8	130	9.7	3.4
6–0 to 6–5	155	8.0	3.8	175	8.6	3.3
6–6 to 6–11	180	6.4	3.8	60	7.2	3.5
7–0 to 7–5	156	5.1	3.6	61	5.8	3.3
7–6 to 7–11	110	4.2	3.4	47	4.6	2.8
8–0 to 8–5	62	3.4	3.1	53	4.2	2.5
8–6 to 8–11	60	2.7	2.8	60	3.0	2.5
9–0 to 9–5	65	2.2	2.5	78	2.8	2.2
9–6 to 9–11	49	1.8	2.2	47	2.3	2.1
10–0 to 10–5	27	1.5	1.8	76	1.9	1.9
10–6 to 10–11	31	1.2	1.5	68	1.8	1.8
11–0 to 11–11				73	1.4	1.4

[1] $N = 1104$; socioeconomic cross section: 98% white, 2% nonwhite.
[2] $N = 975$; socioeconomic cross section: 86% white, 8.5% black, 1% Oriental, 4.5% Mexican-American and Puerto Rican.

Source: From Elizabeth M. Koppitz *The Bender Gestalt Test for Young Children, Vol. 2: Research and Applications 1963–1973*. Published by Allyn and Bacon, MA. Copyright © 1973 by Pearson Education. Reprinted by permission of the publisher.

Table 12.7 Bender indicators of brain injury for children 5 to 10 years of age[1]

Extra or Missing Angles

Figure A Significantly[2] more often in BI at all age levels.

Figure 7 Common in BI and NBI through more frequently in BI at all age levels; *no* BI drew correct angles before age 8.

Figure 8 Common in BI and NBI through age 6, significant[2] for BI thereafter.

Angels for Curves

Figure 6 Common in BI and NBI but significantly[2] more often in BI at all age levels; *all* BI drew angels up to age 7.

Straight Line for Curves

Figure 6 Rate but highly significant[3] for BI when present.

Disproportion of Parts

Figure A Common in BI and NBI through age 6, significant[2] for BI thereafter.

Figure 7 Common in BI and NBI through age 7, significant[2] for BI thereafter.

Substitution of Five Circles for Dots

Figure 1 Present in BI and NBI but significantly[2] more often in BI at all ages.

Figure 3 Present in BI and NBI through age 6, significant[2] for BI thereafter.

Figure 5 Present in BI and NBI through age 8, significant[2] for BI thereafter.

Rotation of Design by 45°

Figures 1, 4, and 8 Highly significant[3] for BI at all age levels.

Figures A and 5 Significant[2] for BI at all age levels.

Figure 7 Present in BI and NBI through age 6, significant[2] for BI thereafter.

Figure 3 Present in BI and NBI through age 7, significant[2] for BI thereafter.

Figure 2 Present in BI and NBI through age 8, significant[2] for BI thereafter.

Failure to Integrate Parts

Figures A and 4 Significant[2] for BI at all age levels.

Figure 6 Rare but significant[2] for BI at all age levels.

Figure 7 Common for BI and NBI through age 6, significant[2] for BI thereafter.

Omission or Addition of Row of Circles

Figure 2 Common in BI and NBI through age 6, highly significant[3] for BI thereafter.

Shape of Design Lost

Figure 3 Present in BI and NBI through age 5, significant[2] for BI thereafter.

Figure 5 Rare and does *not* differentiate between BI and NBI at any age.

Line for Series of Dots

Figures 3 and 5 Rare but highly significant[3] for BI at all age levels.

Perseveration

Figures 1, 2, and 6 Common in BI and NBI through age 7, highly significant[3] for BI thereafter.

[1] BI = brain-injured; NBI = non-brain-injured.

[2] Significant = occurring more often, but not exclusively, in BI group.

[3] Highly significant = occurring almost exclusively in BI group.

Source: From Elizabeth M. Koppitz *The Bender Gestalt Test for Young Children, Volume 1*. Published by Allyn and Bacon, MA. Copyright © 1963 by Pearson Education. Reprinted by permission of the publisher.

4. Rotation of one or more figures 90 degrees or more.
5. Incorrect number of units in three or more figures.
6. Perseveration from figure to figure of one type or more.
7. Tremulous line quality.
8. Lines instead of dots.
9. Drawing a straight line when a curved one is indicated.

Visual-Motor Perception Difficulties Difficulties in visual-motor perception might be caused by emotional factors, developmental delay, CNS complications, or a combination of all of these. Often, it is the task of the clinician to understand both the underlying causes of the visual-motor difficulties and the manner in which these difficulties affect the child. The specific pattern of effects can be noticed by observing the child's behavior during the test as well as the type and severity of errors on the reproductions of the designs. Some children might have primary difficulties with rotations, which might reflect mirror reversals involved with other tasks, such as reading. In contrast, other children might have difficulties in sequencing, which could be suggested by a poorly confused sequence in the reproduction of their Bender Gestalt designs. This qualitative analysis of Bender Gestalt protocols should always be conducted in the context of additional material in the client's history as well as other test data.

Sometimes children have learned to compensate for visual-motor difficulties caused by CNS complications. As a result, their actual Bender Gestalt reproductions might be relatively accurate. This compensation is particularly likely if an injury is not too extensive, there was above-average premorbid intelligence, the location of the lesion is not too critical, and the injury has not been recently acquired. If children have achieved an adequate level of compensation, their actual Bender Gestalt reproductions might be quite accurate. Clinicians can sometimes detect the possible presence of brain damage by becoming sensitized to a wide range of possible compensatory mechanisms. Koppitz (1975) has listed some of these:

- Excessive length of time for completion.
- "Anchoring" designs by placing a finger on them as they attempt to reproduce them.
- Reproducing a design from memory after first glancing at it.
- Checking and rechecking the number of dots, yet still being uncertain regarding the correct number that should be included.
- Rotating either the sheet of paper or the Bender Gestalt card itself as an aid in reproducing the design.
- Designs that are quickly and impulsively drawn and then corrected with extreme difficulty.
- Expressions of dissatisfaction with the poorly reproduced designs followed by repeated efforts to correct them.

It is sometimes useful to attempt to determine whether a child's poor Bender Gestalt reproductions are the result of inadequate reception (difficulty in visual perception) or

inadequate expression (difficulty in reproducing that which might have been accurately perceived). This distinction can sometimes be determined by asking the child to evaluate the accuracy of the drawing he or she has made. Children who feel that their poorly reproduced drawings are accurate most likely have receptive difficulties and possibly difficulties with expression. If they recognize that their drawings were done poorly, this suggests their problem might be primarily expressive. Although they might be aware of the inaccuracy of their drawings, they would be expected to have extreme difficulty in correcting the inaccuracies.

Emotional Indicators The most clearly supported use of the Bender Gestalt is as a screening device for neuropsychological impairment and as an index of developmental delay. However, it has also been used as a projective device to measure various personality functions (Koppitz, 1963, 1975; Reichenber & Raphael, 1992; Rossini & Kaspar, 1987). This usage has been most successful when cutoff scores have been established for various indicators of emotional difficulties. However, often children with visual-motor difficulties resulting from developmental delay also have numerous emotional indicators. Despite this, a significant number of cases have several emotional indicators without necessarily having scores that indicate developmental delay. In these cases, the emotional indicators assess different levels of functioning. As Koppitz (1975) summarizes, "Not all youngsters with poor Developmental Bender Gestalt test scores necessarily have emotional problems, nor do all children with Emotional Indicators on their Bender Gestalt records inevitably show malfunctioning or immaturity in the visual-motor area" (p. 83).

Koppitz (1963, 1975) has listed 12 emotional indicators and developed a scoring manual for 10 of these, which she includes in her 1963 text. Each indicator has specific interpretive hypotheses associated with it. However, these specific hypotheses have not received sufficient empirical support, so any interpretation based on these indicators should be speculative. In contrast, far greater success has been achieved at predicting difficulties such as psychopathology, acting out, and anxiety by using summed totals of indicators (Oas, 1984; Rossini & Kaspar, 1987). Koppitz (1975) recommended using three or more indicators as the cutoff for inferring emotional difficulties. She reports that more than 50% of children with three indicators were emotionally disturbed, and 80% with four or more indicators had serious emotional problems. Any person with three or more indicators should be given a more complete evaluation to determine the nature and extent of possible difficulties. It is likely that a high number of persons with fewer than three indicators may still have significant difficulties and yet might be misclassified as normal. Rossini and Kaspar suggest that one indicator is not uncommon for normal controls; children with adjustment problems often have two to three, and three or more indicators are typical among behavior-disordered children. Thus, they recommend a more conservative approach: Two or more indicators suggest psychopathology. Although the total number of indicators has been used to successfully distinguish psychotic from neurotic levels of psychopathology in children (McConnell, 1967), it has not been successful in discriminating difficulties in nonclinical samples of schoolchildren (M. Gregory, 1977).

Koppitz (1963) originally listed 10 emotional indicators and later expanded these to 12 in her 1975 update of her system. Rossini and Kaspar (1987) found that the following three emotional indicators were given strong support:

1. Confused order.
2. Large size.
3. Box around design.

The following five indicators were found less frequently in Rossini and Kaspar's (1987) control group of normals. Although they did not relate significantly to psychopathology, their less frequent occurrence in the normal controls suggested some possible relationship with emotional difficulty:

4. Expansion.
5. Fine line.
6. Careless overwork (or heavily reinforced lines).
7. Second attempts (without correcting the original).
8. Small size.

The following four indicators did not relate to psychopathology in Rossini and Kaspar's (1987) sample, but have been found to be indicators by other researchers:

9. Wavy line.
10. Dashes for circles.
11. Increased size.
12. Elaboration.

Quite possibly, indicators 4 to 12 may relate to types of pathology other than the adjustment and behavior disorders used in Rossini and Kaspar's (1987) sample of 7- to 10-year-olds. Also, older populations (e.g., adolescents) having emotional problems might be more likely to have a wider variety of emotional indicators. Future research may eventually refine Koppitz's 12 indicators and identify those that have been demonstrated to be the most powerful predictors.

Block Design and Object Assembly

Both Block Design and Object Assembly are sensitive to lowering resulting from any type of organic impairment but are especially sensitive to damage in the right parietal region (Golden, 1979; Lezak, 1995). An advantage of both subtests is that careful behavioral observation can help the practitioner more fully understand a client's deficits. Clients with perceptual difficulties do poorly, primarily because they distort and misperceive the design with a frequent sense of disorientation. These difficulties are more consistent with right parietal lesions. In contrast, patients with left parietal lesions are able to correctly perceive the overall gestalt of the design, but their problem-solving

style may be confused and simplified. Other clients might be able to understand the task and perceive it correctly, but still experience difficulty in actually completing the task. This dissociation between intent and actually being able to make the blocks do what they want is formally referred to as *constructional dyspraxia.* Sometimes clients with a concrete orientation to problem solving do quite poorly on Block Design, because it requires a certain degree of abstraction. In contrast, Object Assembly deals with more concrete objects with the result that these same persons might perform relatively better on it. Additional interpretive details can be found by consulting the relevant sections on these subtests in Chapter 5 and Groth-Marnat (2000a, Chapter 5 in *Neuropsychological Assessment in Clinical Practice: A Guide to Test Interpretation and Integration*).

Human Figure Drawings and Other Free Drawing Tasks

Whereas the preceding tests (Bender Gestalt, Block Design, Object Assembly) are visuoconstructive tasks in a structured situation, free drawing (visuographic) tasks are far less structured. Clients must initiate, organize, and monitor their activity to a greater extent. As such, they add a new dimension to the more structured Bender Gestalt and Wechsler subtest tasks. In addition to the Draw-A-Person, quality of drawings can be assessed with drawings from the House-Tree-Person, drawings of bicycles, clocks, or the more simpler drawings of a square or Greek cross. Formal scoring criteria and norms can be found for clock, bicycle, and house drawings in Lezak (1995, pp. 584–585), Spreen and Strauss (1998, pp. 483–488), and Lacks (2000, pp. 405–406).

MENTAL ACTIVITIES (ATTENTION AND SPEED OF INFORMATION PROCESSING)

The maintenance of an optimum amount of mental activity involves a complex variety of functions related to filtering, selecting, focusing, shifting, and tracking (see Ponsford, 2000). Because there is typically a huge amount of available information to attend to, a person must be able to filter this potential information and attend to only the most relevant sources. Any irrelevant information must be ignored. This filtering, selecting, and focusing process is still not sufficient in and of itself. Unless a person can shift attention, he or she will have difficulty functioning. Attention must strike a balance and be neither overly focused nor too ready to shift. An individual who becomes too focused expresses this symptomatically in perseverations. Such persons then experience difficulty shifting their attention to a new task and are, therefore, likely to continue with a behavior beyond the point in which it is adaptable. Conversely, people who shift their focus too readily express this symptomatically in distractibility.

Because of the complexity and interrelationship with other tasks, attention is quite sensitive to the effects of CNS complications. It is thus one of the most frequently reported disturbances associated with cerebral impairment (Lezak, 1989b). The most basic form of assessment for attentional deficits is through simple reaction time tasks. For example, reaction time has been found to be sensitive to the effects of head trauma (Van Zomeren & Brouwer, 1990), solvent exposure (Groth-Marnat, 1993), and the early

impact of dementia (Teng, Chui, & Saperia, 1990). As attentional tasks become more complex, they become progressively more sensitive to the impact of neuropsychological dysfunction. Thus, not only do the tests in this subsection require simple attention, but also clients must effectively filter out irrelevant stimuli and shift their attention. The Trail Making test requires them to sequence their responses, and, on Trails B, they must shift their attention back and forth from numbers to letters and scan the page while still maintaining the correct sequence of responses. To further assess for attentional difficulties, clinicians can note performances on the WAIS-III/WISC-III attention-related tests as measured by the Working Memory/Freedom from Distractibility Index (Arithmetic, Digit Span, Letter-Number Sequencing) and Processing Speed Index (Digit Symbol-Coding, Symbol Search; see Chapter 5 and Groth-Marnat et al., 2000), for interpretation of these subtests. The result should be a relatively thorough overview of attention and related functions.

Trail Making Test

The Trail Making Test (Army Individual Test, 1944; Reitan & Wolfson, 1993) is an easily administered, widely used test that requires a client to draw lines connecting consecutively numbered circles (Part A; Figure 12.1) followed by a similar task in which they draw lines connecting alternating numbered and lettered circles (Part B; see Figures 12.1 and 12.2). Scores are based on the total time it takes to complete Part A and the total time it takes to complete Part B. It is part of the comprehensive Halstead-Reitan Neuropsychological Test Battery but is frequently used as a component of other comprehensive or screening batteries. It was originally developed by U.S. Army psychologists and is considered to be in the public domain. It can thus be reproduced without obtaining permission. Alternate forms have been developed (Trails C and D; McCracken & Franzen, 1992) that can be used for retesting when it would be important to minimize increases in performance because of practice effects (Dye, 1979).

Trail Making is frequently listed under tests assessing orientation and attention (Lezak, 1995; Groth-Marnat, 2000a). However, it involves a series of skills related to attention including complex scanning, coordination, visuomotor tracking, speed of information processing, and motivation (Gaudino, Geisler, & Squires, 1995). Consistent with this is that Trail Making has loaded most heavily on both "visuospatial sequencing" and "rapid visual search" factors (desRosiers & Kavanagh, 1987). A client completing the task must understand the symbolic importance of the numbers and letters, effectively scan the page, accurately identify the next stimulus to respond to, and perform these functions in a relatively fast manner. Left hemisphere abilities are most likely required to understand and correctly sequence the numbers and letters, and right hemisphere abilities are required to visually scan the page. Doing this in an accurate and quick manner is likely to reflect an intact integration of these abilities. Although Part A has sometimes been used to reflect left hemisphere functioning and Part B to reflect right hemisphere abilities, this lateralization of functions should not be made (Gaudino et al., 1995; Reitan & Wolfson, 1993; Wedding, 1979). In contrast, Trail Making should more accurately be considered to reflect the overall integrity of general brain functioning. However, there is some evidence that patients with dorsolateral

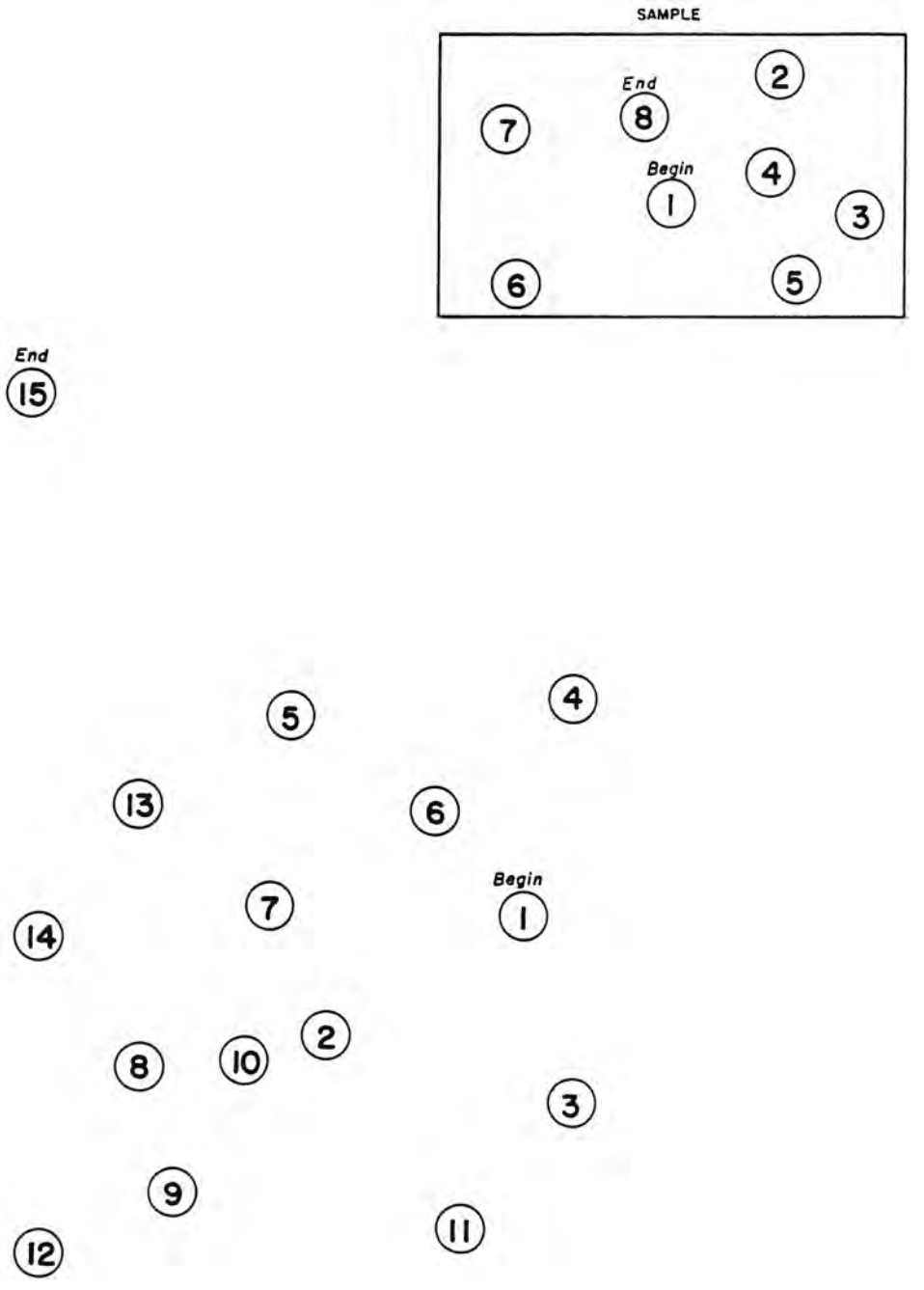


Figure 12.1 Trail Making Part A (abbreviated/child version). Note that Part A and the sample have been reduced, and they are ordinarily on separate $8\frac{1}{2}'' \times 11''$ sheets of paper.

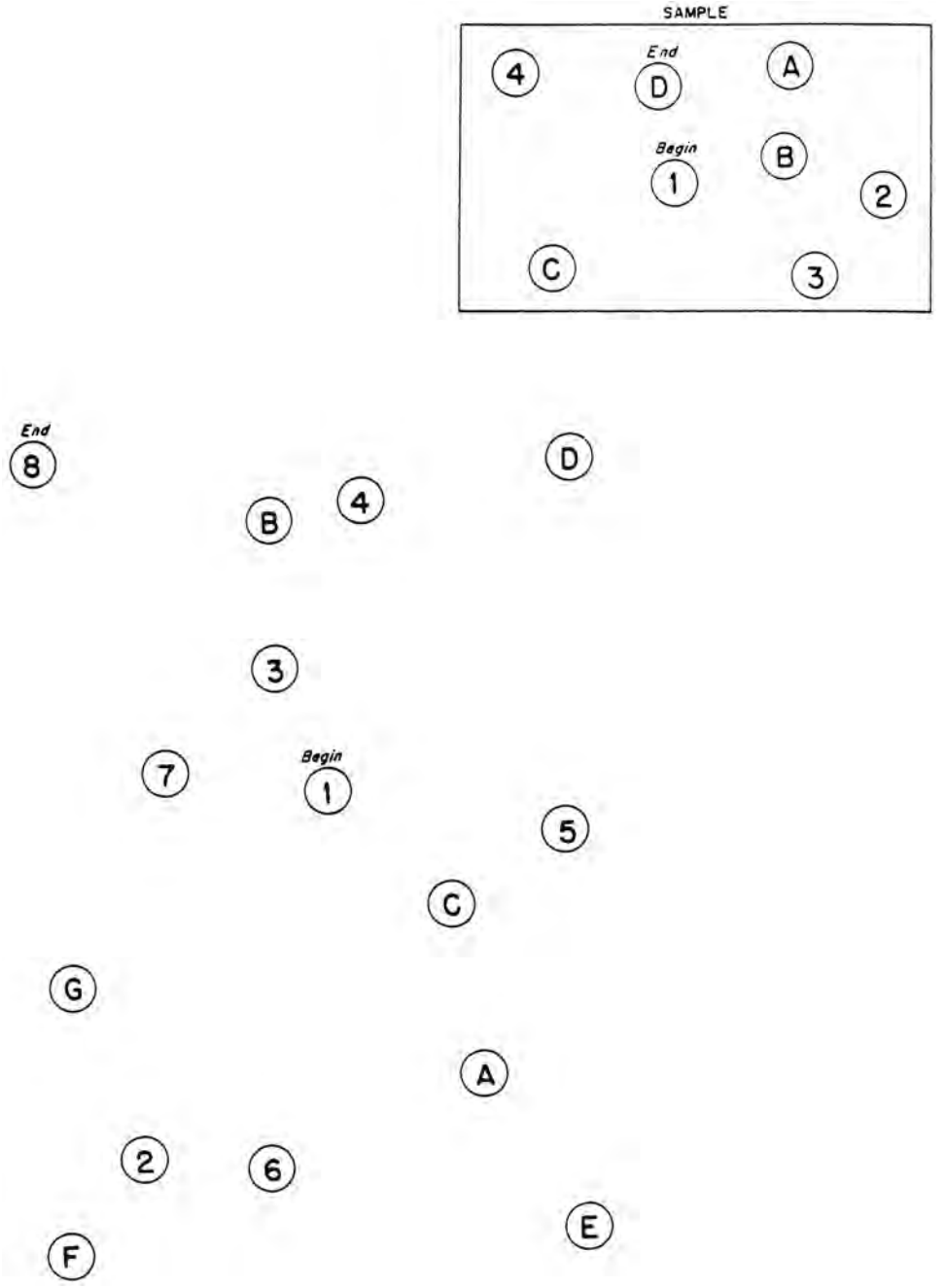


Figure 12.2 Trail Making Part B (abbreviated/child version). Note that Part B and the sample have been reduced, and they are ordinarily on separate 8½″ × 11″ sheets of paper.

frontal lesions (but not other frontal areas) have both slower performances and more errors on Trails B (Stuss et al., 2001). Despite this finding, it should be stressed that Trails B should not be used as a measure of general frontal lobe dysfunction.

Reliability and Validity

Most reports of reliability have been above .60 with some in the .90s (Lezak, 1995; Spreen & Strauss, 1998). Reliability for a group of healthy adults over a 6- to 12-month interval was .78 for Part A and .67 for Part B (Lezak, 1982). However, there is considerable variability among different populations. For example, neuropsychiatric patients having vascular disorders had a test-retest reliability of .94 for Part A (G. Goldstein & Watson, 1989). Similarly, epileptics had test-retest reliabilities over a 6- to 12-month retesting period ranging between .78 and .89 for Part A and .39 and .87 for Part B (Dodrill & Troupin, 1975). The lowest overall reliabilities were found for schizophrenics, who had a reliability of only .36 on Part A (G. Goldstein & Watson, 1989).

Trail Making, like other tests that load heavily on attention, is quite sensitive to the effects of CNS deficits. In particular, Trail Making has been effectively used to detect the early stages of dementia as well as track the progressive decline in abilities during the course of dementia (Botwinick, Storandt, Berg, & Boland, 1988; Rasmusson, Zonderman, Kawas, & Resnick, 1998). There is also slowing of Trail Making performance following head injury with greater slowing reflecting the severity of the injury (Leninger, Gramling, & Farrell, 1990). Performance on Trail Making has also been used to effectively monitor the improvement of patients over time (Stuss, Stethem, Hugenholtz, & Richard, 1989). In addition, significantly slower performances on Trails B have been found for clients with mild hypoxemia (Prigatano, 1983) and chronic toxic encephalopathy (Nilson, Barregard, & Baeckman, 1999). Ecological validity for Trail Making is suggested in that scores were able to predict the extent to which moderately to severely head-injured patients could function independently (Acker & Davis, 1989).

An important consideration with Trail Making is the extent to which it can effectively distinguish between severe psychiatric conditions such as schizophrenia and brain damage. Crockett, Tallman, Hurwitz, and Kozak (1988) found that Trail Making scores could differentiate between psychiatric clients, general medical patients, and patients with neuropsychological impairment. Whereas emotional factors seem to lower Trail Making performance, they rarely lower it as much as brain damage (Gass & Daniel, 1990). Thus, the MMPI, particularly any elevations on Scales 6 (Paranoia), 7 (Psychasthenia), and 8 (Schizophrenia), can be used to determine the extent to which emotional factors might be involved in poor Trail Making performance. However, some authors have cautioned that, in a number of cases, Trail Making performance has had difficulty distinguishing between psychiatric and brain-damaged patients (Norton, 1978). This is conceptually consistent with the view that attentional difficulties can be disrupted not only by brain damage, but also by conditions such as anxiety and depression.

Administration

The task for Trail Making requires clients to accurately connect the correct sequence of numbers and letters by drawing a line from one to the next as quickly as possible. The client should be seated comfortably at a table and supplied with a pencil. The correct sequence of the test (practice samples followed by Trails A and then Trails B; see

Figures 12.1 and 12.2) is then given to them. If a client makes an error, the examiner needs to detect this error as quickly as possible and then place the client's pencil back on the preceding (correct) circle as soon as possible. If a client intended to touch a circle but did not quite do so, this should not be counted as an error, but the client should be asked to actually touch or, preferably, enter the circle during future trials. The following instructions have been adapted from Reitan and Wolfson (1993, pp. 279–288) and Spreen and Strauss (1998, pp. 533–535):

Practice Sample for Part A Place the practice sample for Part A (see Figure 12.1) directly in front of the client. Point to the sheet and say: **"On this page are some numbers. Begin at number 1"** (point to circle 1), **"and draw a line from 1 to 2"** (point to circle 2), **"2 to 3, 3 to 4, and so on in order, until you reach the end"** (point to the circle designated as END). **"Draw the lines as fast as you can. Do not lift the pencil from the paper. Ready, begin!"**

If the client makes an error on practice sample A (or Part A itself), you can provide the following explanations depending on the type of error made:

- **"You started with the wrong circle. This is where you start"** (point to circle 1).
- **"You skipped this circle"** (point to the one skipped). **"You should go from number 1 to 2, 2 to 3"** (point to each circle), **"and so on until you reach the circle marked END"** (point).
- **"Please keep the pencil on the paper, and continue right on to the next circle."**

If the client still cannot complete the practice sample, take the eraser of the pencil, place his or her hand around it, and guide the person through the trail with the eraser on the page. Then turn the pencil around with the point facing downward and say, **"Now you try it. Put your pencil point down. Remember begin at number one and draw a line from one to two, two to three, three to four"** (point to each of the numbers), **"and so on, in order until you reach the circle marked END. Ready, Begin!"**

When clients have difficulty, repeat the instructions until they can demonstrate that they understand what is required. If they still cannot complete or cannot adequately understand the task, discontinue the test. However, if the client completes the task correctly, immediately go on to Part A of the test.

Say, **"On this page are numbers from 1 to 25. Do this the same way. Begin at number 1 and draw a line from 1 to 2, 2 to 3, 3 to 4"** (point to each circle with the designated number in it) **"and so on, in order, until you reach the end"** (point to the circle designated END). **"Remember to work as fast as you can. Ready, begin!"**

Begin timing as soon as the instructions have been given. If a client makes an error, point out the error immediately, guide the person's pencil back to the prior (and correct) circle, and have him or her continue the test from this point. Do not discontinue timing while you are correcting the client. The error is taken into account only in that it increases the time it takes to complete the entire Part A task. After completing Part A, note the final time and say, **"That's fine. Now we'll try another one."** Then proceed immediately to the practice sample for Part B.

Practice Sample for Part B Place the practice sample page (see Figure 12.2) directly in front of the client and say, **"On this page are some numbers and letters. Begin at 1"** (point to 1) **"and draw a line from 1 to A"** (point to circle A), **"A to 2"** (point to circle 2), **"2 to B"** (point to circle B), **"B to 3"** (point to circle 3), **"3 to C"** (point to circle C), **"and so on, in order, until you reach the end"** (point to the circle indicated as END).

"Remember, first you have a number" (point to circle 1) **"then a letter"** (point to circle A), **"then a number"** (point to 2), **"then a letter"** (point to B), **"and so on. Draw the line as fast as you can. Ready, begin!"**

If the client makes a mistake on practice sample B (or Part B itself), you can, based on the type of error, provide the following instructions:

- **"You started with the wrong circle. This is where you start"** (point to circle 1).
- **"You skipped this circle"** (point to the circle that was skipped). **"You should go from circle 1 to A, A to 2, 2 to B, B to 3"** (point to each of the correct circles) **"and so on until you reach the circle marked END"** (point to the circle indicated as END).
- **"You went only as far as this circle"** (point to the last circle the person went to). **"You should have gone to the circle marked END"** (point to the circle indicated as END).
- **"Please keep the pencil on the paper, and go right on to the next circle."**

If the client still cannot complete the practice sample, take the eraser of the pencil and, with the person's hand around it, guide him or her through the trail with the eraser on the page. Then turn the pencil around with the point facing downward and say, **"Now you try it. Remember you begin at number 1 and draw a line from 1 to A, A to 2, 2 to B, B to 3, and so on until you reach the circle marked END"** (point to each one of the circles to clarify). **"Ready, begin!"** If the client completes the practice sample or can demonstrate adequate understanding of the procedure, immediately go on to Part B.

Say, **"Now let's try this one. On this page are both numbers and letters. Do this the same way. Begin at number 1 and draw a line from 1 to A, A to 2, 2 to B, B to 3, 3 to C, and so on in order"** (point to each one of the circles as it is being referred to), **"until you reach the end. Remember, first you have a number, then a letter, then a number, then a letter, and so on, until you reach the end. Do not skip around, but go from one circle to the next in the proper order. Draw the lines as fast as you can. Do you have any questions? Ready, begin!"**

Begin timing as soon as the directions for Part A and B have been completed. Parts A and B are scored separately with the scoring ending as soon as the client completes each of the two parts. Scoring for each part is the total number of seconds it takes to complete each task. Any errors should be noted and written down as an additional interpretive consideration.

Interpretation

While scores on Trail Making can generally be considered to assess abilities associated with attention, it is often difficult to determine more specifically which related skills are involved. The clearest factor loadings indicate the importance of speed of visual search and visuospatial sequencing (desRosiers & Kavanagh, 1987). Additional areas to consider are level of motivation, depression, poor coordination, or conceptual confusion. Often behavioral observations and types of errors can be quite useful in refining the meanings of low or high scores. Some errors, particularly among head-injured patients, might reflect impulsivity because they jump ahead to an incorrect number/letter (Lezak, 1989b). Perseverative errors might be reflected in difficulty

alternating between numbers and letters in Trails B. Scores can be used clinically to make inferences related to the presence of scanning and tracking problems.

The interpretive guidelines indicate that low scores reflect the patient's difficulty dealing with more than one stimulus at a time and maintaining a flexible mental orientation. Because flexible thinking is important in everyday activities, it suggests the person has difficulty working effectively in employment settings and, if scores are sufficiently low, living independently. Particular difficulties might develop when the person is attempting to perform tasks requiring divided attention (performing more than one task at the same time). This is likely to be particularly true for low Trails B scores. In general, because Trails B is a more complex task, it is both more sensitive to impairment and more useful in making inferences regarding a client's level of functioning.

Low Trail Making scores further suggest that a client has a more general difficulty with executive functions related to initiating, inhibiting, sequencing, and monitoring his or her behavior. Relevant behavioral observations to note are difficulty initiating behavior unless specifically directed to do so, impulsiveness as suggested by beginning behavior without being requested to do so, or perseverations. The clinician should also interview family members to determine the extent of coaching or prompting that the client requires and the extent to which he or she can function independently.

Reitan and Wolfson (1993) provide a general classification of scores for Trails A with the normal range from 0 to 39; mildly impaired, 40 to 51; and moderate to severe impairment, 52 or more. Normal Trails B performances range between 0 and 85; mild impairment, 86 to 120; and moderate to severe impairment is 121 or more. However, because Trail Making is influenced by age, education, and intelligence, most authors recommend using scores corrected for age and education (Heaton et al., 1991; Spreen & Strauss, 1998, pp. 539–542). Because age has the most influence on performance, Table 12.8 presents age-related norms developed by several authors and compiled by Spreen and Strauss (1991, 1998). This set of norms has the advantage of conveniently presenting percentiles so that low scores (especially under the 10th percentile) can be

Table 12.8 Trails test: Time in seconds (on Parts A and B) for normal control subjects at different age levels

	15–20 Years (n = 108)		20–39 Years (n = 275)		40–49 Years (n = 138)		50–59 Years (n = 130)		60–69 Years (n = 120)		70–79 Years (n = 90)	
Percentile	A	B	A	B	A	B	A	B	A	B	A	B
90	15	26	21	45	18	30	23	55	26	62	33	79
75	19	37	24	55	23	52	29	71	30	83	54	122
50	23	47	26	65	30	78	35	80	35	95	70	180
25	30	59	34	85	38	102	57	128	63	142	98	210
10	38	70	45	98	59	126	77	162	85	174	161	350

Source: Data extrapolated from Davies, 1968, based on a representative British (Liverpool) sample; Fromm-Auch and Yeudall, 1983; and Kennedy, 1981. From *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary,* p. 326, by Otfried Spreen and Esther Strauss. Copyright © 1991 by Oxford University Press, Inc. Reprinted by permission.

easily determined. However, education correlates .19 for Trails A and .33 for Trails B (R. A. Bornstein, 1985); therefore, level of education should be noted and taken into account when making interpretations.

Digit Span, Arithmetic, Digit Symbol-Coding, Symbol Search, Letter-Number Sequencing

The Working Memory/Freedom from Distractibility Index on the WAIS-III/WISC-III is composed of Digit Span, Arithmetic, Letter-Number Sequencing, and, sometimes, Digit Symbol-Coding. Each of these subtests requires subjects to pay careful attention to the task. Digit Symbol-Coding and Symbol Search also assess speed of information processing. They can, therefore, be used as measures of a person's attentional abilities. Because these tasks also assess a person's ability to remember, process, and learn (and attention and memory are closely related), they have also been included in the section on Learning and Memory. More extensive interpretive guidelines, including relevant neuropsychological issues, can be found in Chapter 5 and Groth-Marnat et al. (2000, Chapter 5).

One option for extending the sensitivity of Digit Symbol is to administer the quite similar Symbol Digit Modalities Test (SDMT; A. Smith, 1982). However, SDMT is different in that the symbols are printed on the test form and the client must write in the numbers (the opposite occurs for Digit Symbol-Coding). In addition, SDMT includes not only a written presentation of the task, but also an oral administration. The test is quite sensitive to a wide range of cerebral dysfunction, is particularly effective in discriminating depression from dementia (A. Smith, 1982), and is sensitive to the effects of head injuries (Ponsford, 2000; Ponsford & Kinsella, 1992). The reason for using it in addition to Digit Symbol-Coding is that SDMT appears to be a more sensitive indicator of cerebral dysfunction as indicated by instances in which Digit Symbol-Coding was normal, whereas SDMT was in the impaired range (A. Smith, 1982).

MEMORY AND LEARNING

The types and procedures of memory and learning are complex (see Baddeley, Wilson, & Watts, 1995; Helmes, 2000). Aspects of these processes might include sensory memory, short-term memory, rehearsal, long-term memory, consolidation, recall, recognition, and forgetting. In addition, memory and learning can be divided into two major subsystems: declarative memory, which refers to learning about information, objects, and events; and procedural or implicit memory, which refers to automatic, habitual responses. Each of these subdivisions has somewhat different anatomical structures. Additional useful subdivisions of memory are verbal versus spatial, automatic versus effortful, and semantic versus episodic. Studies of brain-lesioned patients indicate that memory can be further divided into extremely specific subareas based on functions such as sensory modality (verbal, tactile, auditory, etc.), type of material (verbal, motor skill, etc.), or content of information (numbers, letters, pictures, names, faces, etc.; Baddeley et al., 1995). For the practitioner, providing a truly comprehensive evaluation of memory functions is a daunting task.

Fortunately, a more limited number of memory domains can usually provide practitioners with an overview of the general intactness of memory. These include (a) the extent to which the subject can acquire and retain new material, (b) how quickly material is forgotten, (c) the extent to which competing information interferes with learning, (d) the degree of specificity or generality of the deficit, and (e) the stability or fluctuation of the deficits over time (Walsh, 1994). Ideally, these domains should include measurements of both visual and verbal material.

To more fully assess the complex and multifactorial structure of learning and memory, a number of relatively comprehensive memory batteries have been developed. Among the oldest, and certainly the most frequently used (Camara et al., 2000), is the Wechsler Memory Scale (WMS; Wechsler, 1945, 1974), which was revised in 1987 (Wechsler Memory Scale-Revised; Wechsler, 1987) and again in 1997 (Wechsler Memory Scale-III; Wechsler, 1997b). The WMS-III has eight primary indexes, assesses both verbal and visual-spatial functions, includes a delayed recall component, and takes approximately 45 minutes to administer (see Chapter 6). Additional relatively comprehensive batteries include the Rivermead Behavioral Memory Test (B. Wilson, Cockburn, & Baddeley, 1985), Memory Assessment Scales (J. M. Williams, 1991), Wide Range Assessment of Memory and Learning (W. Adams & Sheslow, 1990), and the Denman Neuropsychology Memory Scale (Denman, 1984).

One important distinction is between attention versus memory and learning. In some ways, this distinction is inappropriate because attention is a prerequisite for learning to occur. A person who is easily distracted does not effectively learn and remember relevant information or events. Attention is, therefore, closely linked to learning. However, in other ways they do represent distinct functions. In particular, it is important to distinguish whether a person is capable of learning but is easily distracted, or whether, even under circumstances in which the person fully attends to a task, he or she still cannot learn very efficiently. This sometimes happens when clients state that they have a memory problem, but, despite their symptom description, they perform learning and memory tasks quite well under the ideal circumstances that often characterize assessment procedures. In contrast, real-world situations frequently mean that they need to exclude a number of distractions and carry on two or more activities simultaneously. Under these conditions, they might have distinct difficulties dividing their attention and, therefore, might not be able to learn and remember particularly effectively. Interviewing them regarding situations in which they do versus don't remember effectively might help the practitioner to understand this issue better. In addition, their test performances would be expected to be lower on tasks that load more heavily on attention (Trails B, Arithmetic, Digit Span, serial sevens or serial threes) than those that are more pure tests of learning (Rey Auditory Verbal Learning Test, repeating paragraphs/stories, Bender Gestalt memory).

A good beginning place to assess memory is in the interview. Details regarding basic information such as personal, family, educational, and employment history can be pursued. Interviewers might request dates when the client began or finished employment or education, parents' or children's birthdays, or details related to medical history. Some of this information might be compared with more objective sources to determine its accuracy. In addition, behavioral observations such as pauses, expressions of uncertainty, or confusion might suggest difficulties with retrieval.

Current research consistently indicates that there is a mild to moderate relationship between memory impairment and depression. An extensive meta-analysis by Burt et al. (1995) found that memory impairment was most clearly associated with inpatients (versus outpatients) and mixed bipolar and unipolar patients (versus purely unipolar). In addition, negative affective information was more likely to be remembered accurately than material with a positive or neutral emotional tone. However, memory impairments were also present among populations of schizophrenics and mixed groups of psychiatric patients but not among patients diagnosed with either anxiety disorders or substance abuse. Interestingly, the association between memory and depression was stronger among younger than older persons. This is probably because early onset depression is likely to be more severe and younger persons have a greater amount of memory to lose (greater "ceiling" and "floor") than older persons (narrower range between ceiling and floor). Despite these findings, it should also be stressed that the link between memory impairment (and other forms of neuropsychological functioning) and depression is typically of quite a small magnitude (Burt et al., 1995; Sherman et al., 2000). For example, dementia typically accounts for a far larger proportion of the variance in neuropsychological functioning than depression.

The tests recommended in this subsection provide a useful slice of memory functions relevant to populations with CNS deficits. The WAIS-III/WISC-III subtests of Digit Symbol-Coding (incidental learning), Information, Digit Span, and Letter-Number Sequencing (WAIS-III only) include potentially valuable information related to learning and memory. However, Digit Span and Letter-Number Sequencing are primarily attentional tasks rather than pure learning tests. In addition, the Rey Auditory Verbal Learning Test is a relatively brief, well-researched, frequently used, individually administered test that assesses short-term verbal memory, the ability of the client to learn new material, the extent to which interference disrupts learning, and the ability to recognize information that might have been previously learned. As the name suggests, however, it is verbally oriented. To include at least some visual-spatial memory assessment, the memory version of the Bender Gestalt is recommended. If a more thorough assessment of visual-spatial memory is required, clinicians might consider the Benton Visual Motor Retention Test (Benton, 1974), the visual portions of the WMS-III (Faces, Family Pictures, Visual Reproduction), or the Rey-Osterrith Complex Figure Test (Meyers & Meyers, 1996; Osterrith, 1944; Rey, 1941; Spreen & Strauss, 1991).

Rey Auditory Verbal Learning Test

The Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964; M. Schmidt, 1996; Spreen & Strauss, 1998) uses a simple format in which the client is asked to remember a list of 15 unrelated words (List A) repeated over five different trials. The client is then presented with another list of 15 unrelated words (List B), which serves to potentially interfere with previous learning, followed by a request to recall as many of the words from the original list as possible. After a 30-minute delay, the client is again asked to recall words from the original list (List A), following which he or she is asked to recognize as many words as possible in a list that includes words from the original list. As a result, a wide diversity of functions can be assessed. These include short-term auditory-verbal

memory, rate of learning, learning strategies, retroactive and proactive interference, presence of confabulation or confusion in memory processes, retention of information, and differences between learning and retrieval. The entire procedure takes 10 to 15 minutes.

One past difficulty with the RAVLT was the lack of a manual with standard administration and scoring procedures. This has now been amended with the publication of a manual along with meta-norms by M. Schmidt (1996). Several authors have developed alternative lists for examiners wishing to conduct follow-up evaluations and avoid the difficulties of practice effects. These lists are available for the original list (List A) and the interference list (List B), as well as the longer lists used for the recognition task. They are summarized in Helmes (2000, p. 319), Lezak (1995, pp. 439, 441), and Spreen and Strauss (1998, pp. 330–331). In addition, D. M. Shapiro and Harrison (1990) have developed four alternative sets of lists that have been found to be equivalent to the original RAVLT.

For testing children, a simpler Children's Auditory Verbal Learning Test (CAVLT; J. Talley, 1990) that uses the same format as the RAVLT is commercially available with a manual, scoring keys, and set of children's norms (Talley, 1990). Another commercially available variation of the RAVLT is the California Verbal Learning Test (CVLT; Delis et al., 1987). Instead of the RAVLT's list of unrelated words, the CVLT uses more conceptually consistent items that might be found in a typical shopping list. While scoring can be quite complex, there is a computer program to help with calculating some of the more elaborate scores. The test is well normed and has been found sensitive to important clinical areas such as the early effects of Alzheimer's disease and the differential diagnosis between various types of dementias (Massman, Delis, Butters, Dupont, & Gillin, 1992).

Reliability and Validity

Test-retest reliability of the RAVLT over a one-year interval was a somewhat moderate .55 (Snow, Tierney, Zorzitto, Fisher, & Reid, 1988). The highest reliability was .70 for the total number of words recalled for the five trials of List A. In contrast, the lowest reliability was .38 for recall of List B (Snow et al., 1988). The importance of using alternate forms is highlighted by the finding that practice effects for the same form over a 6- to 12-month retesting period were small but significant (1 to 2 words per trial; Crawford, Stewart, & Moore, 1989; Lezak, 1982). In contrast, no differences were found when alternate forms were used (Crawford et al., 1989).

Consistent with expectations, patients with left hemisphere damage have been found to have lower performances than those with damage to the right hemisphere (Ivnik, Sharbrough, & Laws, 1988; Miceli, Caltagirone, Gainotti, Masullo, & Silveri, 1981). In addition, the RAVLT has been found to be sensitive to the effects of different memory disorders. Heavy drinkers scored poorly on the RAVLT even if they did not have signs of neurologically related disease (M. Jackson, Fox, Waugh, & Tuck, 1987). As would be expected, Korsakoff's patients did consistently poorly on each of the five trials but, when presented with a recognition format after each of the first five trials, their performances increased (Janowsky, Shimamura, & Squire, 1989). In a similar administration format, frontal lobe patients had poor recall over the five trials of List A, but when

asked to recognize which words had been on the list using a recognition list, their performances were near normal (Janowsky et al., 1989). This suggests that their difficulties were mainly due to retrieving, organizing, and keeping track of the answers related to any potentially learned material rather than to their not having learned it. Finally, patients in the early stages of Alzheimer's disease showed a slow learning curve in which they recalled an average of only 6 of the 15 words following the fifth trial of List A (Mitrushina, Satz, & Van Gorp, 1989). They also had far more words intruding between the A and B lists than other diagnostic groups (Bigler, Rosa, Schultz, Hall, & Harris, 1989; Mitrushina et al., 1989). When the disease progressed to a moderate level, the number of words recalled for the fifth trial dropped to approximately 5, and for severe cases only an average of 2.6 words were recalled (Tierney et al., 1994).

Further support comes from correlations ranging from .50 to .65 between RAVLT factor groupings and other learning instruments (MaCartney-Filgate & Vriezen, 1988). A factor analytic study with normals indicated that the tests measured the functions of acquisition, storage, and retrieval (Vakil & Blachstein, 1993).

Administration

Word lists and a scoring format can be found in Appendix K on page 698. The following set of instructions is derived from guidelines and verbatim instructions provided primarily by Lezak (1983, 1995, pp. 438–440) and, to a lesser extent, by Spreen and Strauss (1998, pp. 327–328). The reliance on Lezak's instructions is emphasized because most research studies specify that they have relied on her guidelines. For Trial 1, examiners should state:

> **"I am going to read a list of words. Listen carefully, for when I stop, you are to repeat back as many words as you can remember. It doesn't matter in what order you repeat them. Just try to remember as many words as you can."**

Each word in List A should be read with a one-second interval between the words. No feedback should be given regarding whether they have given correct responses, repeated words, or included words not on the list. However, clients can be encouraged for their efforts on this, as well as additional trials. The order in which the words are recalled can be indicated by numbering them on the scoring sheet.

When the client indicates that he or she is unable to think of additional words, Trial 2 for List A can be given with the following instructions:

> **"Now I am going to read the same list again, and once again, when I stop, I want you to tell me as many words as you can remember, *including words you said the first time.* It doesn't matter in what order you say them. Just say as many words as you can remember, whether you said them before or not."**

Trials 3 to 5 are also given again using the words from List A and following the preceding instructions. Again, a one-second interval between each word should be provided. Each of these trials should be initiated only after the client has indicated he or she cannot recall any words from the list that has just been read.

After completing Trial 5 for List A, the examiner should then begin Trial 6 (and using List B) by saying:

"Now I am going to read a second list of words. This time, again, you are to say back as many words of this second list as you can remember. Again, the order in which you say the words does not matter. Just try to remember as many as you can."

Immediately after the client indicates that he or she is unable to recall additional words from List B, begin Trial 7 by requesting that he or she recall as many words as possible from the previous list (List A). Note that List A is not read to the person before this request.

After an interval of 30 minutes, begin Trial 8 by asking the client to repeat as many words as possible from the original list of words (List A). Again, the list should not be read to the person. Directly after the client indicates the inability to think of additional words, begin Trial 9 by providing him or her with the recognition list of 50 words listed at the bottom of Appendix K. This list contains all the words in both Lists A and B (indicated as either A or B). In addition, words that are semantically similar to words on Lists A or B (coded as *SA* or *SB,* respectively) are included plus words that are phonemically similar to words on lists A and B (coded as *PA* and *PB,* respectively). In some cases, words are both semantically and phonemically similar and are indicated as either *SPA* or *SPB.* Practitioners are encouraged to develop a separate listing of the recognition list but exclude the codes (*SA, SB,* etc.). This can then be shown to the client without the risk of their deciphering the meanings associated with the codes.

Any incorrect responses can be abbreviated using the following designations: *R* for words that are repeated, *RC* if repeated but then self-corrected, *RQ* if the client has repeated the word but indicates being unsure if this has occurred or not, and *E* for words that are not on the list. This coding can provide clinicians with a quick review of the types of errors made.

A wide number of different scoring categories have been developed (see Geffen, Moar, O'Hanlon, Clark, & Geffen, 1990; Ivnik et al., 1990). To keep the scoring and interpretation procedures manageable, the following list describes the most useful and frequently used scores:

1. *Immediate Memory.* Score for Trial 1; based on the total number of words correctly recalled from word List A immediately following the first trial (scoring for the next four trials is also based on the total number of words correctly recalled after being presented with word List A on successive occasions).
2. *Best Learning.* Score for Trial 5; based on the total number of words correctly recalled from word List A immediately following the fifth trial.
3. *Total Learning.* Sum of the total number of words correctly recalled for Trials 1 to 5.
4. *Proactive Interference.* Score for Trial 6; based on the total number of words correctly recalled from List B immediately following the sixth trial.
5. *Retroactive Interference.* Score for Trial 7; based on the total number of words correctly recalled from List A (directly following Trial 6 and without having word List A repeated).
6. *Delayed Recall.* Score for Trial 8; based on the total number of words correctly recalled following a 30-minute delay and without having word List A repeated.

7. *Recognition.* Score for Trial 9; based on the total number of words correctly recalled when presented with the recognition list.

Norms for each of the seven scoring categories are included in Appendix L on page 700. However, clinicians skilled in the use of the RAVLT can increase the variety of information by noting additional features such as the number and types of words that intrude between Lists A and B or assessing the relative extent to which words at the beginning of a list are more likely to be recalled than those at the end of a list (primacy versus recency effect).

Interpretation

The norms summarized in Appendix L (and adapted from M. Schmidt's, 1996 meta-norms) have been organized around the preceding scoring categories. One representative norm is that for *immediate memory;* healthy adult males between ages 13 and 79 correctly recalled between 6.8 and 5.5 words depending on their age grouping. This same group correctly recalled between 13.1 and 10.3 words for the *best learning* trial (total number of words recalled for Trial 5 of List A), and between 11.8 and 8.1 words were recalled for *retroactive interference* (total number of words correctly recalled from List A following the administration of the Interference Trial for List B). To provide a contrast, moderately impaired Alzheimer's disease patients recalled only 2.7 ($SD = 2.1$) words for *immediate recall,* 5.3 ($SD = 3.4$) words for best learning, and 1.8 ($SD = 3.0$) words for *retroactive interference* (Tierney et al., 1994).

Normative studies have found RAVLT performance gradually declines with age, with a more pronounced decline after the age of 70 (Geffen et al., 1990; Savage & Gouvier, 1992). Education and intelligence have been found to have relatively little influence on RAVLT performance (J. Ryan, Rosenberg, & Mittenberg, 1984; Savage & Gouvier, 1992; Weins, McMinn, & Crossen, 1988). In contrast, females have been found to perform slightly better than males for some of the scoring categories (Geffen et al., 1990). However, this greater female performance rarely exceeded more than an average of one word per category.

Hypotheses related to client functioning can be derived from considering level of performance on the different scoring categories listed in Appendix K (and L). For example, immediate or short-term memory can be assessed by noting scores on the *immediate memory* scoring category. Similarly, the degree to which new learning interferes with past learning can be inferred by noting the level of performance on the scoring category for *retroactive interference.* However, practitioners must use caution when making these inferences. Although the RAVLT has been successful in both identifying patients with memory disorders and even distinguishing between these disorders, insufficient validity studies have been done to investigate the validity of more specific interpretations. For example, it is tempting to interpret low scores on Trial 7 as indicating that the person is particularly susceptible to retroactive interference. However, this interpretation has not been systematically investigated. In addition, the reliability of some of the scores, particularly recall for List B, is quite low (Snow et al., 1988). This means that interpretations based on scores derived from recall of List B (Trial 6/Proactive Interference) should be made with considerable caution.

A client's learning curve is a useful index. Normal persons under the age of 70 usually remember between 6 and 8 words on Trial 1, and this slowly increases to between 12 and

14 words for the final trial (Trial 5). In contrast, if little or no learning has occurred, few additional words are recalled by Trial 5. Moderately impaired Alzheimer's and Parkinson patients recall an average of 2.7 and 2.9 words for Trial 1, and their performances increased to an average of only 5.3 words each by Trial 5 (Tierney et al., 1994). Another potentially useful pattern is that sometimes brain-damaged patients can do as well as normal healthy persons on the first trial but have difficulty increasing their recall much beyond this original performance. A further pattern is that normal persons typically show a primacy effect in which they recall more words from the first part of the list as well as a recency effect in which they recall more words from the end of the list (those words heard most recently). Clients with memory impairments who are struggling with the material are more likely to have a recency effect without a corresponding primacy effect.

A useful comparison is to note the difference in performance between the first recall for List A (Trial 1; *immediate memory*) and the (first) recall for List B (Trial 6; *proactive interference*). Usually, the number of words recalled for Trial 6 is only about one word fewer than for Trial 1. If Trial 6 is lower than Trial 1 by two to three words or more, it suggests that the person has difficulty "freeing up" abilities to attend to new learning (proactive interference). This hypothesis is further strengthened if words from List A "intrude" into the words recalled from List B (as requested on Trial 6).

One qualitative difference between patients with Alzheimer's disease versus those with depression is a greater number of intrusions between lists of words for the Alzheimer's patients (Burt et al., 1995; Ober, Koss, Friedland, & Delis, 1985). This represents a general trend among Alzheimer's patients to be liberal in their responses, resulting in more intrusion errors as well as more false positives on recognition-type tasks (Gainotti & Marra, 1994). In contrast, depressed patients tend to be conservative in their responses. A further distinction is that depressed patients tend to do relatively poorly on immediate memory tasks but are able to consolidate the information such that their delayed recall is relatively good. In contrast, Alzheimer's patients do more poorly on delayed than on immediate recognition, indicating primary difficulties in consolidating and retaining the information (Burt et al., 1995). Further differences between these two groups are that depressed persons have more variable performances, whereas Alzheimer's patients are more consistent in performance; and the level of concentration among depressed persons is adequate, whereas it is quite likely to be deficient among Alzheimer's patients (see desRosiers, 1992).

Similar to the difficulty with intrusions between Lists A and B, interference because of the new task of having to learn List B (in Trial 6) can be noted by comparing the difference between performance on Trial 7 (recall of the previously learned List A as measured by score on *retroactive interference*) and performance on Trial 5 (recall of List A after five trials as measured on score for *best learning*). Usually the number of words recalled on Trial 7 is approximately one word fewer for a normal performance. A lowering of two to four words or more suggests that the task of learning List B has interfered with the previously learned List A.

One way to organize patterns of scores is to group them around the following RAVLT memory factors of acquisition, storage, and retrieval (Vakil & Blachstein, 1993):

- Acquisition
  Immediate memory.
  Proactive interference.

- Retrieval
  Delayed recall.
  Retroactive interference.
  Best learning.
- Storage
- Recognition

The preceding clustering should provide a more theoretically and psychometrically sound analysis of memory than merely using individual scoring categories. In addition, it can help specify the various strengths and weaknesses of a particular client's memory functions. For example, the previous information on depression and memory indicates that depressed patients would be slow to acquire the information because their attention and motivation is typically poor. In contrast, Alzheimer's patients would also have some difficulties in acquiring the information, but their storage would be particularly poor because they would make a considerable number of errors in recognizing previous words from the different lists.

Digit Span, Letter-Number Sequencing, Digit Symbol-Coding, Information

An important source of memory functions can be found throughout various sections of the WAIS-III/WISC-III. Digit Span and Letter-Number Sequencing (WAIS-III only) involve the assessment of both attention (and thus its inclusion in the Mental Activities section) and short-term memory. Digit Symbol-Coding similarly requires intact attention and short-term memory, but also requires the person to effectively rote-learn simple nonverbal material. The relative importance of speed/attention versus memory can be parceled out by using the optional incidental memory and copy procedures. In contrast to the short-term emphasis of Digit Span, Digit Symbol-Coding, and Letter-Number Sequencing, Information allows an examiner to assess the extent of remote memory. It is usually difficult, however, to distinguish between lack of knowledge and retrieval difficulties. One way to assess this is to note whether the client's level of knowledge is clearly discrepant with his or her level of education (e.g., a person with a university education who doesn't know the capital of a major country). Another strategy is based on the relative ease by which material can be recognized as opposed to recalled. Thus, clients might be presented with multiple-choice questions (available as part of the WAIS-R NI; E. Kaplan et al., 1991 and WISC-III PI; E. Kaplan et al., 1999) in which a greater recognition result suggests that they once knew the material but had a difficult time retrieving it during the standard recall instructions. The hypothesis that retrieval is a difficulty for a client should be supported by a corresponding relatively low score on the RAVLT Retrieval Factor (composed of scores on Delayed Recall, Retroactive Interference, and Best Learning).

Bender Gestalt (Memory Administration)

A very general idea of visual memory can be obtained by requesting the client to draw as many of the Bender Gestalt designs as possible directly after completing the last Bender

Gestalt design with the standard procedure (see discussion, pp. 533–534). A more detailed assessment of visual memory can also be obtained through additional more extensive procedures (visual memory portions of the Wechsler Memory Scale-III, Rey-Osterrith Complex Figure, Benton Visual Motor Retention).

VERBAL FUNCTIONS AND ACADEMIC SKILLS

Disturbances of verbal functions are frequently associated with brain damage, particularly when the damage is to the left hemisphere. As a result, any review of neuropsychological functions needs to assess verbal functions as well as the academic skills that are frequently associated with these verbal abilities. The most common disturbances are the aphasias (impaired speech, writing, or understanding spoken or written language) and problems with speech production. These disorders can involve extremely diverse difficulties including poor articulation, loss of verbal fluency, word-finding difficulty, poor repetition of words or sentences, loss of grammar and syntax, misspoken words (paraphasias), poor auditory comprehension, reading difficulties, and impaired writing (Goodglass & Kaplan, 1983).

As a result of the variety of these disorders, a neuropsychological screening and initial review as recommended in this chapter can assess only a relatively small number of them. For a full assessment of aphasic and related disorders, several comprehensive batteries are available, including the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983), Communicative Abilities in Daily Living (Holland, 1980), and the Multilingual Aphasia Examination (Benton & Hamsher, 1989). In contrast to these formal, comprehensive batteries, Lezak (1995) recommends an informal and general clinical review of six major functions as follows:

1. *Spontaneous Speech.* Observe how clients initiate, articulate, and organize their speech.
2. *Speech Repetition.* Ask clients to repeat words, phrases, and sentences. In particular, this might include repeating difficult words such as *Massachusetts* or *Methodist Episcopal* (see Reitan & Wolfson's, 1993, Aphasia Screening Test) to assess for disorders of articulation.
3. *Speech Comprehension.* Request that they answer simple questions (e.g., Is a ball square?) or obey simple commands (e.g., point to specific objects, put their hands on their chins).
4. *Naming.* Ask clients to name common objects, colors, letters, and actions.
5. *Reading.* Have clients read aloud; for comprehension, have them explain what they have read.
6. *Writing.* Request that the subject copy, write to dictation, and compose a sentence.

Choosing which tasks to give can be based on the hypotheses derived from other information related to the client (symptom checklist, medical history, etc.). The relative difficulty of the tasks can similarly be tailored to additional information regarding client symptoms and behaviors. For example, it would be neither necessary nor

appropriate to request that a client with mild deficits obey simple commands or name common objects.

The following recommended tests emphasize five of the Wechsler scales that load most strongly on verbal and academic abilities: Information, Comprehension, Similarities, Vocabulary, Arithmetic. Each of these contributes somewhat different types of information (see discussions in Chapter 5). The Controlled Oral Word Association test (COWA; Benton & Hamsher, 1989; Spreen & Strauss, 1998) was selected because word fluency and word finding are frequent complaints among brain-damaged populations (Benson, 1993). Available evidence indicates that the COWA can effectively identify these complaints (Benton, 1968; Murdoch, Chenery, Wilks, & Boyle, 1987; R. Parks et al., 1988). In addition, it does not require any equipment, is easily administered, short, frequently used, and has good psychometric properties including excellent normative data (see Benton & Hamsher, 1983; Lezak, 1995; Spreen & Strauss, 1998).

The Boston Naming Test has similar assets and might also be considered for screening and assessment. Somewhat similar to the COWA, it assesses disturbances in word finding and word naming. Clients are asked to look at pictures of objects and provide their names (tree, abacus, etc.). Research has indicated that it can effectively identify word naming and finding difficulties (Margolin, Pate, Friedrich, & Elia, 1990; Morris et al., 1989; Spreen & Strauss, 1998; Storandt et al., 1986) and, because of its sensitivity to the early effects of dementia, a shortened version has been included in the CERAD battery (Morris et al., 1989). Clinically, naming difficulties might be noted during the course of an interview or somewhat more formally by noting difficulties in naming different objects in Picture Completion. These observations might suggest a more thorough assessment with the Boston Naming Test.

The most frequently used educational achievement battery in clinical neuropsychology (Camara et al., 2000) is the Wide Range Achievement Test, which is now available in its third edition (WRAT-III; Jastak & Wilkinson, 1994). The battery is easy to administer, covers a wide range of ages (12 to 75 years), and provides scores for spelling, reading, and arithmetic. These can each be conveniently portrayed as school grade equivalents, standard scores, or percentiles. However, it assesses a somewhat narrow range of abilities in these domains and thus should be used only as a crude screening instrument. An increasingly popular, more in-depth assessment of achievement can be obtained from the Wechsler Individual Achievement Test (WIAT; Wechsler, 1992). It has the advantages of having norms that are linked with the WISC-III and a screen for 10 to 18 year olds that can guide the selection of additional WIAT subtests.

Controlled Oral Word Association

The Controlled Oral Word Association test (COWA; Benton & Hamsher, 1989; Spreen & Strauss, 1991) simply requests that a client say as many words as possible beginning with a certain letter and within a certain time limit. The COWA has also been variously referred to as *Word Fluency* or, more frequently, simply the *FAS Test* because the letters *F, A,* and *S* are the ones most commonly used. A similar and much earlier version was developed by Thurstone (1938), who used written word production, and Newcomb (1969), who recommended alternating between colors and birds. It has also been included in the much longer Neurosensory Center Comprehensive Examination for

Aphasia (Spreen & Benton, 1977), as well as the Iowa Screening Battery for Mental Decline (Eslinger, Damasio, & Benton, 1984) and the Multilingual Aphasia Examination (Benton & Hamsher, 1989).

The client must freely produce words, yet also monitor his or her previous responses. A factor analysis indicated that it loaded most strongly (.62) on "abstract mental operations," which makes it similar to tasks such as oral spelling, mental calculations, and digit span-type tests (Snow et al., 1988). Somewhat differently, des Rosiers and Kavanagh (1987) found that it loaded primarily on a "verbal knowledge" factor, thereby making it closer to tests such as Vocabulary and those that contribute to a person's overall Verbal IQ.

Reliability and Validity

Test-retest reliability for adults over a 19- to 42-day retesting interval was .88 (desRosiers & Kavanagh, 1987), whereas it dropped to .70 for older adults with a much longer retesting interval of one year (Snow et al., 1988). Because of the clarity of scoring (number of words recalled), interscorer reliability was almost perfect.

COWA scores are sensitive to a variety of CNS impairments, and the patterns of performance conform to what is known regarding brain function. Lesions to the frontal (especially left) lobes have been found to most noticeably produce word fluency difficulties that are reflected in lower COWA performance (Miceli et al., 1981; Perret, 1974). The lowest overall word fluency scores have been with bilateral frontal lesions (Benton, 1968). Dementing patients have also shown lowerings in word fluency (Murdoch et al., 1987), although COWA scores have not been able to effectively differentiate between patients with dementia and those who are elderly and depressed (R. Hart, Kwentus, Taylor, & Hamer, 1988). In addition, patients with right hemisphere damage (excluding right frontal) have not been found to be impaired on COWA (Cavalli, De Renzi, Faglioni, & Vitale, 1981).

Administration

Instructions provided by Spreen and Strauss (1991) are as follows:

> **I will say a letter of the alphabet. Then I want you to give me as many words that begin with that letter as quickly as you can. For instance, if I say "B," you might give me "bad," "battle," "bed." I do not want you to use words that are proper names such as "Boston," "Bob," or "Brylcreem." Also, do not use the same word again with a different ending such as "eat" and "eating." Any questions?** (pause) **Begin when I say the letter. The first letter is "F." Go ahead.**

As soon as the instructions have been given, the examiner should begin timing. One minute should be allowed for each of the different trials for the letters *F, A,* and *S.* Indicate when each of the three trials is completed by saying "Fine" or "Good." If, after initially producing only a few words, the client becomes silent, encourage the person to continue trying to produce additional words. This might require providing a brief synopsis of the instructions. Write down all words in the order in which they were produced. The score is the sum of all admissible words across the three trials. Inadmissible words (i.e., proper nouns, wrong words, repetitions) should not be counted for the final score.

Interpretation

While scoring COWA is straightforward, performance is considerably influenced by education, age, and, to a lesser extent, gender. Specialized norms need to be consulted (see Tables 12.9 and 12.10) and interpretations made taking these demographics into consideration.

High scores (relative to age and education) suggest that the person has good verbal knowledge (check Vocabulary and Similarities) and is competent in dealing with abstract mental operations (check Arithmetic, Digit Span, and Letter-Number Sequencing). Low scores suggest the person has difficulties in these areas. However, the meaning of a low score can be refined by considering the patterns of other test scores along with relevant behavioral observations. If the client demonstrates difficulties with naming objects, such as struggling with some of the visual stimuli on Picture Completion, word finding and naming (agnosia) might be the primary impairment. It is possible for verbal fluency difficulties and visual naming deficits to occur separately (Benson, 1993). On the other hand, a client who has low scores on tests of attention (Arithmetic, Digit Span, Letter-Number Sequencing, Trail Making, and, possibly, Digit Symbol-Coding) may be having difficulties with distractibility and may be unable to maintain a focus on the task. Alternatively, mental inflexibility might be the major reason for lowered performance on these tests. This might be assessed by noting the presence of any perseverations or lack of flexibility as seen on poor performances with Trails B.

Table 12.9 Data for the sample stratified by age and education

				FAS		
					Percentiles	
Age	*N*	Mean Age	Education	Mean (*SD*)	5th	10th
By Age						
25–34	309	31.0	16.1	45.7	26	30.5
		(2.6)	(2.2)	(12.7)		
35–44	290	39.3	16.4	46.1	26	29
		(2.9)	(2.3)	(12.6)		
45–54	97	48.5	16.7	45.9	25	29
		(2.6)	(2.6)	(12.3)		
By Education						
<College	229	36.1	13.7	41.7	23	26
		(7.4)	(1.2)	(11.6)		
College	202	35.6	16.0	46.2	28	31
		(7.2)	(0.0)	(12.3)		
>College	302	38.4	18.6	49.0	29	32
		(7.8)	(1.3)	(12.4)		

Note: The table presents data for the total number of words generated on FAS test within 1 minute.
Source: Adapted from Selnes et al., 1991.

Table 12.10 Letter fluency performance for cognitively normal participants according to age, gender, and educational level

	FAS (Letter Fluency)		
Group	*M*	*SD*	*N*
Age Group (Years)			
65–74	24.0	12.4	139
75–84	25.8	11.5	343
85+	24.0	10.8	146
Gender			
Male	23.2	12.1	258
Female	26.2	11.0	370
Education (Years)			
0–6	16.2	6.9	140
7–9	23.7	9.9	170
10–12	27.0	10.2	202
13+	34.2	12.6	115
Sample	25.0	11.6	628

Source: Adapted from Crossley, D'Arcy, and Rawson, 1997.

Vocabulary, Information, Comprehension, Similarities, Arithmetic

Both Vocabulary and Information are heavily dependent on past learning, require good verbal abilities, and are relatively resistant to deterioration (Lezak, 1995). Thus, they provide good measures of an individual's general verbal abilities. In addition, both these subtests are associated with activation of the left temporal lobe (Chase et al., 1984). Qualitative responses, especially on Vocabulary, can be quite useful and include clang responses, idiosyncratic associations, or confabulations. Comprehension measures a person's common sense, judgment, and practical reasoning. As such, it is not as dependent on past education as Vocabulary and Information. Because the questions are open-ended and require the client to exercise judgment, the qualitative responses to the different items can be quite useful in understanding the client's reasoning processes. In contrast to the more crystallized abilities of Vocabulary and Information, Similarities is more of a fluid task and is also more sensitive to left hemisphere lesions (Warrington, James, & Maciejewski, 1986), especially to the left frontal (Rzechorzek, 1979) and bilateral frontal regions (Rao, 1990). It is also quite sensitive to the presence and progression of dementia (R. Hart et al., 1988).

Whereas the Wechsler subtests primarily assess verbal abilities, Arithmetic is somewhat different because it also has a very strong memory and attentional component as well as mathematical and verbal components. Thus, it comprises a number of different interacting abilities. Low scores do not necessarily mean that the client has low mathematical abilities, especially if other tasks loading heavily on attention and concentration

are also low. However, if a person's verbal and attentional abilities are intact, scores are much more likely to indicate the level of a person's mathematical abilities based partially on his or her past academic background. More detailed discussions of each of the preceding subtests can be found in Chapter 5 and Groth-Marnat et al. (2000, Chapter 5).

TESTS OF MOTOR ABILITY

Measurements of motor abilities can frequently be used to assess for subtle motor impairment as well as make inferences regarding lateralization of lesions. The expected difference is that the dominant hand will be 10% stronger and faster than the nondominant hand. Differences of 20% or more have frequently been used to infer lesions contralateral to the side of weakness/slowing (Reitan & Wolfson, 1993). While these inferences tend to hold true when several different measures are used (R. A. Bornstein, 1986; L. Thompson, Heaton, Mathews, & Grant, 1987), measurements by a single test need to be treated with caution. Equal levels of dexterity as measured by finger-tapping speed occur relatively frequently among normal populations. Conversely, differences between finger-tapping speeds occur relatively frequently such that, if a 20% discrepancy is used for finger-tapping speed among left-handed persons, 18% are misclassified as having lateralized hemisphere dysfunction (Thompson et al., 1987). Localization (including lateralization) also needs to be treated cautiously because of the variety of locations that might produce right-left motor discrepancies. Because a test such as finger tapping requires planning and initiation of behavior (executive functions), it might be impaired because of anterior frontal or even subcortical involvement. Alternatively, impairment might be caused by damage in the motor strip or perhaps by coordinative dysfunction from cerebellar damage. Thus, the following test(s) should be used for hypothesis generation rather than diagnosis.

Finger Tapping Test

The Finger Tapping Test (FTT) is an essential component of the Halstead-Reitan Neuropsychological Test Battery (Reitan & Wolfson, 1993). It was originally referred to as the *Finger Oscillation Test* and, in some sources, is still referred to by that name. While some coordination is required, it is primarily a test of simple motor speed. The client is asked to initially tap his or her dominant index finger for five consecutive 10-second trials. The procedure is then repeated for the nondominant index finger. Performances are measured on a recording device (available from Reitan Neuropsychology Laboratory, Lafayette Instruments, or Western Psychological Services). The score is simply the average number of taps in a 10-second interval. The two average scores (for dominant and nondominant fingers/hands) are compared with each other to see if there are wide discrepancies; normative comparisons can also be made.

Reliability and Validity

Test-retest reliability for normal controls over a 10-week interval was quite high (male $r = .94$, female $r = .86$; D. Gill, Reddon, Stefanyk, & Hans, 1986). A longer retesting interval of six months for another group of normals still revealed relatively

good reliabilities of .71 for the dominant hand and .76 for the nondominant hand (Ruff & Parker, 1993). A mixture of clinical samples (vascular disease, head trauma, alcoholics, schizophrenics) for a mean two-year retesting interval similarly showed good reliabilities ranging between .64 and .87 (G. Goldstein & Watson, 1989).

One of the main rationales for the FTT is that it should reflect CNS dysfunction contralateral to the finger with slowed tapping speed. This has been generally supported (G. Brown, Spicer, & Robertson, 1989; Finlayson & Reitan, 1980; Reitan & Wolfson, 1993). However, lateralized damage might still occur without being reflected in lowered speed, particularly if the damage is in the posterior cortex (Reitan & Wolfson, 1993). In addition, finger-tapping speed has been found to correlate with a surprising number of complex functions including awareness of deficits, outcome from rehabilitation, and the development and gradual loss of rote verbal learning throughout the lifespan (Prigatano, 1999).

Administration

The client is instructed to tap as rapidly as he or she can with the index finger on a small lever. Recording begins with the dominant finger/hand for a 10-second trial. This is repeated over five trials with a mandatory rest of two to three minutes after the third trial. This procedure is then repeated for the nondominant hand. Never alternate trials between dominant and nondominant hands; all five trials should be given first for the dominant, and then for the nondominant hand. The following instructions adapted from Reitan and Wolfson (1993, p. 232) should be given to the client:

> **"Now we are going to do a test to see how fast you can tap. We will use this little key here"** (indicate the key to the client) **"and I want you to tap just as fast as you can, using the forefinger of your right"** (or left, depending on which is dominant) **"hand.**
>
> **When you tap, be sure to use a finger movement; do not move your whole hand or arm. When you tap this key, you will have to remember to let the key come all the way up and click each time, or else the number on the dial won't change."** (Demonstrate how the lever works and tap it for five or six seconds.)
>
> **"Now you move the board to a comfortable position for your hand and try it for practice."**
>
> After a brief period of practice, say:
>
> **"That was fine. Remember to tap as rapidly as you possibly can."**
>
> Make sure that the client understands the instructions and then say:
>
> **"Do you have any questions? . . . All right. Ready—Go!"** At the end of the 10-second trial say, **"Stop!"** Then repeat this procedure for another four trials. After the client has completed the procedures with the dominant hand, repeat for the nondominant hand.

Several additional considerations are relevant. First, examiners must take care to start the stopwatch as soon as the client begins tapping. Similarly, the timing must cease as soon as the 10-second interval is finished. If clients continue to tap after the examiner says "stop," the number of taps should be noted and subtracted from the score. Second, be sure that only the index finger does the tapping. Any extraneous movements of the hand or arms should be discouraged. This can be best accomplished by instructing the client to keep the heel of the hand on the base of the tapping board or table.

Interpretation

Averaged scores for dominant and nondominant hands can be interpreted using two different approaches. The first is to assess the discrepancy between performance on the dominant versus nondominant hands. Usually, there is a 10% faster performance for the dominant as opposed to the nondominant hand. If the discrepancy appears considerably wider than expected, brain damage contralateral to the lowered hand is suspected. Reitan and Wolfson (1993, p. 387) provide the following categories based on dividing the nondominant hand by the dominant hand and subtracting this from 1.0:

.12 (or less)–.18	normal
.19–.26	mild to moderate impairment
.27 (or more)	severe impairment

Diffuse impairment may cause a general lowering in both hands that would make a lateralized effect unlikely to occur. This may mean that it would be most appropriate to note the relative degree of lowering without reference to lateralized differences. To interpret this second strategy, Bornstein, Paniak, and O'Brien (1987) recommend the following cutoff scores for impaired performances:

males	dominant hand ≤ 32
	nondominant hand ≤ 31
females	dominant hand ≤ 26
	nondominant hand ≤ 21

A caution relevant to both approaches is that some clients have developed considerable abilities on one side of the body (e.g., musicians who have well-developed nondominant hand strength and coordination), and examiners should consider this in their interpretations.

EXECUTIVE FUNCTIONS

Executive functions involve a person's ability to effectively regulate and direct self-behavior. These functions can be subdivided into volition, planning, purposive action, and effective performance (Lezak, 1995; Sbordone, 2000b). For example, patients experiencing significant executive impairments might exist in a semivegetative state in which they rarely initiate much activity although other cognitive abilities might be quite intact. Other patients with executive difficulty may have little awareness of their impact on others and thus are unable to effectively direct or regulate their social behavior. While frontal lobe damage is most typically implicated with executive deficits, damage to subcortical, especially thalamic, regions or the more diffuse damage caused by anoxia or organic solvents can also produce executive impairment (see Sbordone, 2000b).

Despite the importance of executive abilities, they can be overlooked during formal psychological assessment, partially because executive functions can be impaired even

though other cognitive functions appear quite intact. As a result, a clinician might look at cognitive test scores such as a composite IQ and erroneously conclude that a patient has made a good or even full recovery. There are even anecdotal reports of patients' IQs actually increasing after frontal lobe damage although they became quite impaired because of a loss of executive abilities. A further reason for failure to assess executive functions is that, until recently, no tests were designed to accomplish this purpose. The early tests that have claimed to assess frontal abilities, such as the Austin Mazes, have been questioned regarding their ability to accurately perform this task (Bowden & Smith, 1994). More recently, however, the Behavioral Assessment of the Dysecutive Syndrome (B. Wilson, Alderman, Burgess Emslie, & Evans, 1999), Frontal Lobe Personality Scale (Grace, Stout, & Malloy, 1999), and the Delis-Kaplan Executive Function System (Delis, Kaplan, & Kramer, 1999) have been developed as comprehensive procedures to measure a wide range of executive abilities. Finally, because much formal assessment is a structured situation in which the examiner directs the patient to do certain things, the patient's ability to self-initiate might be overlooked. This presents examiners with the dilemma that they must "structure" an unstructured situation in which patients can demonstrate the extent, style, and manner in which they would initiate, develop, plan, and monitor their own behavior. A final assessment issue is that, frequently, depression can produce some of the same behaviors (i.e., apathy, flat affect, lack of direction) that occur with executive loss stemming from brain damage. A clinician might, therefore, erroneously conclude that the executive dysfunctions were the result of depression rather than brain damage (or vice versa).

Because of these concerns, strategies to assess executive functions are through various combinations of interview, behavioral observations, and brief informal clinical tests (see H. Hall & Sbordone, 1993; Sbordone, 2000b). Interviews with patients might focus on their articulation of future goals along with their descriptions of recreational activities. Typically, patients with executive difficulties provide little detail about these areas. If they do provide detail, it may be primarily based on reciting their goals and activities before the injury. This means that interviewers need to establish what their present activities and goals are and, in particular, what they have done recently to pursue these goals. Interviewers might also establish the extent to which they can realistically pursue these goals, anticipate and plan relevant activities, develop alternative plans, and give direction to actually putting these plans into action. Because poor executive functions are frequently accompanied by lack of awareness, it might be essential to interview family members who have had a chance to observe the patient on a daily basis. Thus, the client's descriptions can be compared with more objective external descriptions.

In the actual examination itself, various types of behavior can provide information. Does the patient initiate and direct any activity, or does he or she tend to be relatively passive? Are there unusual social behaviors (e.g., poor grooming, discussion of irrelevant tangents, inappropriate jokes) that suggest poor awareness of his or her social impact? The examiner must determine whether such behaviors developed postinjury or were premorbid characteristics. Planning abilities might be estimated based on how well such patients organize their human figure drawings, blocks on Block Design, Bender Gestalt drawings, or their stories on the TAT. Perseverations suggest poor mental flexibility and difficulty monitoring their behavior and are a component of executive

functions; the patient may make too many dots on the Bender Gestalt, have difficulty making mental shifts on Trails B, or find it difficult to understand changes in test stimuli (e.g., slow to understand the requirements of the RAVLT). Because poor executive functions also include difficulty attending to stimuli while simultaneously performing other tasks, low scores on the Wechsler Freedom from Distractibility factor or Bannatyne's Sequencing might also reflect poor executive abilities (Wielkiewicz, 1990).

A number of informal clinical tests might also help to determine possible executive impairments. For example, the patient might be asked to continue the pattern of a drawing that has various repetitive but alternating small shapes (three circles, two squares, one triangle) and then to repeat this sequence several times (see Goldberg & Bilder, 1987). A similar "chain of command" type test is having the patient tap the desk with his or her fist, then tap it with the palm and then repeat this pattern several times. A slightly more complicated task might be as follows: The examiner taps his or her foot once, then the patient taps a foot twice. Alternatively, the examiner may tap a foot twice, while then the patient is instructed to tap once (see Lezak, 1995). None of these procedures have formal scoring; instead, the examiner must determine, based on observation, whether the patient had relative difficulty with all or any of the activities. Although no single strategy in this section is sufficient to identify executive impairments, collectively, the strategies will help to ensure that this critical domain of functioning is included in a client's assessment.

EMOTIONAL STATUS AND LEVEL OF ADJUSTMENT

While measures of cognitive and behavioral abilities are important, a client's emotional status and relative level of adjustment are also of considerable relevance (Gass, 2000; Knight & Godfrey, 1996). This information is useful for at least three types of situations. First, clinicians might be trying to decide whether abnormal cognitive test results are primarily from CNS involvement or emotional factors. If emotional functioning was relatively normal, but the individual still had cognitive deficits, this more strongly implicates CNS involvement. On the other hand, if a client is quite depressed, the depression, not organic factors, might be the primary reason for a symptom such as slowed information processing. Second, a clinician might need to know the extent to which emotional reactions are complicating organic impairment. A client with organically based confusion is likely to have this further exacerbated by reactions such as depression. Third, predictions often need to be made related to a person's overall level of functioning. While level of cognitive deficit is useful, personality and emotional factors have often been found to be better predictors of psychosocial adjustment and rehabilitation outcome (Fordyce, Roueche, & Prigatano, 1983; Heaton et al., 1978).

A wide variety of emotional and personality domains are relevant to neuropsychological impairment. On one end of the spectrum are the negative reactions of anxiety, depression, irritability, emotional lability, suspiciousness, and aggression. At the opposite end are euphoria, emotional flatness, placidity, and naiveté (Prigatano, 1987). Other important areas might be level of self-awareness, interpersonal insensitivity, behavioral rigidity, empathy, and self-centeredness. A common cluster of changes associated with frontal lobe damage is limited self-awareness, impulsivity, concreteness, and poor social

awareness (Stuss et al., 1992). The presence or absence of each of these symptoms has importance for diagnosis, treatment planning, short- and long-range predictions, and feedback to the client and their families.

The assessment of personality and adjustment can be accomplished through a variety of different strategies. The interview might include contrasting the client's descriptions of difficulties with those provided by friends and family. This can help to determine whether, and the extent to which, a client overestimates or denies any deficits (Prigatano, 1992). This also provides an opportunity to obtain descriptions of a client's premorbid characteristics as well as the extent and quality of the person's social supports. Rating instruments designed for use specifically with neuropsychological populations include Patient Competency Rating (Prigatano, 1986), Frontal Lobe Personality Scale (Grace et al., 1999), Neurobehavioral Rating Scale (H. Levin et al., 1987), and the Neuropsychology Behavior and Affect Profile (L. Nelson et al., 1989). In addition, traditional scales including the MMPI/MMPI-2 and BDI/BDI-II have been extensively used. The MMPI/MMPI-2 in particular has the advantage that it provides an index of both stable or trait-type characteristics (social nonconformity, shyness) and more changeable symptom features (depression, anxiety).

MMPI/MMPI-2/MMPI-A and Neuropsychological Impairment

Possible roles for the MMPI/MMPI-2/MMPI-A among neuropsychologically impaired populations are to differentiate between persons who are organically versus nonorganically impaired and to provide information related to persons with known organic complaints. In differential diagnosis, the MMPI/MMPI-2 can often be effectively used to exclude/include complicating psychiatric disturbances. For example, a client who presents with a variety of subtle neurological signs but also has a pronounced conversion *V* on the MMPI/MMPI-2 is more likely to have his or her symptoms based on a somatoform disorder as opposed to organic factors. This hypothesis is further strengthened if medical records indicate that the client characteristically produces similar symptoms when under stress.

The MMPI/MMP1-2 has not been particularly successful in producing either an "organic" profile or in localizing known lesions (Farr & Martin, 1988; Gass & Ansley, 1995; Lezak, 1995) even though some code types, such as 28/82, occur more frequently among brain-damaged groups. They do not occur frequently enough, however, to be diagnostic of CNS involvement. One study found that the 123-code type (neurotic triad) was the most frequent to occur but only did so in 11.2% of a brain-damaged sample (Wooten, 1983). Theoretically and clinically, this makes sense because brain damage is not a unitary phenomenon. It is heterogeneous and is likely to be considerably influenced by a particular client's premorbid personality and style of coping. As a result, a wide diversity of profiles would be expected to (and do) occur among brain-damaged groups.

Despite this heterogeneity, there are individual items that seem to be endorsed more frequently among persons with CNS complications. Many of the items relate to symptoms such as fatigue, weakness, sensorimotor symptoms, distractibility, and paralysis and are more likely to occur among brain-damaged populations. Thus, some of the scales might be inflated not so much because of psychiatric disturbance, but because of

organically based neuropsychological impairment. As a result, clinicians may incorrectly infer, or at least exaggerate, the extent of psychiatric disturbance among brain-damaged populations. To compensate for this, a variety of correction factors has been developed for various brain-damaged populations. These include patients with closed head injury (Gass, 1991), cerebrovascular disease (Gass, 1992), and multiple sclerosis (Meyerink, Reitan, & Selz, 1988). Instructions for MMPI-2 correction scoring, along with interpretive strategies for closed head injury and cerebrovascular disease, are included in Gass (2000, pp. 457–435). Most of these items are derived from Scales 1, 2, 3, and 8; therefore, unless these correction factors are used, clinicians should be somewhat cautious about making interpretations based on elevations on these scales. This caution should be balanced by being aware of and sensitive to the subgroup of brain-damaged persons who have poor awareness of their psychiatric and neuropsychological deficits. Given their minimization of difficulties, they might produce quite normal MMPI/MMPI-2/MMPI-A profiles despite significant pathology.

In summary, the MMPI/MMPI-2 has usually not been effective in diagnosing the presence or location of brain damage, although it can be useful in assessing the current adjustment of brain-damaged persons. This might result from psychopathology that preceded the brain damage, organically based personality changes resulting from the injury, a person's individual reaction to the damage, or a combination. Thus, with the preceding guidelines and cautions, clinicians are referred to Chapter 6 and Gass (2000) for more specific interpretations of MMPI/MMP1-2/MMPI-A profiles.

Beck Depression Inventory (BDI/BDI-II)

The BDI/BDI-II can provide a brief, easily administered assessment of a client's level of depression (see Chapter 13). When working with neuropsychological populations, there is some indication that stroke patients may have somewhat elevated scores because of endorsing items with somatic content (e.g., item 20 related to somatic preoccupation; W. Gordon, Hibbard, Egelkos, & Riley, 1991). This might reflect the physiological effects of the stroke rather than actual psychiatric complications. In contrast, B. Levin, Llabre, and Weiner (1988) demonstrated that Parkinson patients are more likely to have endorsed the BDI somatic items because of endorsing actual psychiatric disturbance rather than to the somatic impact of Parkinson's disease. However, the BDI has relatively fewer somatically related items and more items related to cognitive and affective domains than a scale such as the Hamilton Rating Scale for Depression (C. Brown, Schulberg, & Madonia, 1995). As a result, the BDI would be preferable as scores would be less likely to be incorrectly inflated because of medical conditions. For assessing depression among elderly persons, the Geriatric Depression Inventory (Brink et al., 1982) may be preferable because it was standardized and validated on this population and has relatively fewer items related to somatic concerns (Olin, Schneider, Eaton, Zemansky, & Pollock, 1992).

RECOMMENDED READING

Groth-Marnat, G. (Ed.). (2000). *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration.* New York: Wiley.

Kaufman, D. M. (2001). *Clinical neurology for psychiatrists* (5th ed.). New York: Saunders.

Lacks, P. (1999). *Bender-Gestalt screening for brain dysfunction* (2nd ed.). New York: Wiley.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Snyder, P. J., & Nussbaum, P. D. (1998). *Clinical neuropsychology: A pocket handbook for assessment.* Washington, DC: American Psychological Association.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Oxford University Press.

Vanderploeg, R. D. (Ed.). (2001). *Clinicians guide to neuropsychological assessment* (2nd ed.). Mahwah, NJ: Erlbaum.

Walsh, K., & Darby, D. (1999). *Neuropsychology: A clinical approach* (4th ed.). Edinburgh, England: Churchill Livingstone.

Chapter 13

BRIEF INSTRUMENTS FOR TREATMENT PLANNING, MONITORING, AND OUTCOME ASSESSMENT

Since the early 1990s, there has been an increasing demand for brief, symptom-focused instruments to assist in the delivery of mental health services. One of the major reasons for this is managed care's emphasis on cost containment and documenting treatment efficacy. Most managed care organizations also have a rather narrow symptom-oriented focus on treatment that is quite consistent with the content of many brief clinical instruments. In an effort to contain costs, many managed care organizations have also reduced the reimbursement for testing services such that a full test battery is no longer an option for most practitioners (Eisman et al., 2000). Some managed care administrators have even indicated their belief that psychologists in the past have overused formal testing as a means of padding their incomes. The result has been a reduction in many, if not most, assessment services in many mental health care contexts (Eisman et al., 2000; C. Piotrowski, 1999). The assessment services that are valued by managed care are those that use brief instruments to plan, monitor, and evaluate the impact of interventions.

Another factor emphasizing the importance of brief, focused tests has been the extensive and continually expanding research on the outcomes of mental health interventions. Pretest and posttest measures have proliferated to the extent that there are now a multitude of options to choose from (see Antony & Barlow, 2002; Bufka, Crawford, & Levitt, 2002; Maruish, 1999, 2002). It is almost a given that training and research clinics monitor their work through tests that typically take less than 10 or 15 minutes to complete. Given the models and procedures present in the research arena, managed care organizations are also expecting clinicians to demonstrate that, indeed, the interventions they are implementing are effective (Callaghan, 2001).

The role of brief instruments has also expanded in parallel with the dramatic increase in the areas psychologists have become involved in (Stout, 1997; Stout & Cook, 1999). This has included diverse roles such as prevention, treatment planning, clinical outcomes management, evaluation of psychoactive medication, risk management, malingering, uncovering undiagnosed psychopathology, assessment of chronic pain, geriatric assessment, and behavioral dentistry. Although the potential for a full battery has decreased, these additional areas that are quite likely to use brief instruments have certainly increased.

One of the major challenges confronting psychologists is demonstrating the financial efficacy of their services. There is considerable evidence that psychosocial interventions

are cost effective in psychotherapy as well as general health care (Groth-Marnat & Edkins, 1996). For example, it has long been known that somatizing patients are high overusers of the medical-surgical system. Significant cost savings can be realized by extracting them from the costly (and relatively ineffective) medical-surgical area and into the mental health area where they can receive brief, targeted psychotherapy (Cummings, 1991). Unfortunately, there has been little research into the potential cost savings for assessment services. Rational guidelines suggest that assessment is most likely to demonstrate financial efficacy in the areas of risk management, linking assessment and treatment, using computer-assisted assessment, targeting problems most likely to result in cost savings, use of time-efficient instruments, and focusing on domains of greatest relevance to treatment planning and outcome assessment (Groth-Marnat, 1999).

SELECTING BRIEF INSTRUMENTS

Before selecting a brief screening instrument, there should be some consideration given to who will administer and interpret it. The majority of instruments (and those included in this chapter) are self-report measures. These have the advantages of reducing clinician time and obtaining the client's own perception of their difficulties. However, they also have the potential for bias by the client's perceptions and are potentially subject to under- or overreporting. Other instruments are administered by professional psychologists or allied health professionals such as primary care physicians, nurses, or clinical social workers (see Bufka et al., 2002). On other occasions, a significant other person in the patient's life, such as a parent or spouse, completes the instrument.

The ability of an instrument to assist in planning and outcome assessment is particularly relevant for selecting brief instruments. First, they should not take longer than 15 minutes to complete (and preferably less time). In addition, they should typically be directly relevant to treatment planning and outcome assessment. In contrast, a full battery approach before psychotherapy often provides a large amount of descriptive information, but most of this information is not directly applicable to treatment planning. A further frequent essential quality is that they be useful for screening purposes. For example, the Beck Depression Inventory (BDI-II) can be used for each of the preceding purposes. A physician might use it to detect the possible presence of depression, a psychologist might administer it to determine the baseline level of severity for a client, and this baseline could then be used to determine the effectiveness of interventions targeted to treat the depression. A managed care company would be particularly interested in this process to monitor quality control over the treatments they reimburse.

In addition to time efficiency and relevance to treatment planning and outcome evaluation, brief instruments should also be relevant to various target groups (F. Newman et al., 1999). For example, specialized variations on the BDI have been developed for children and geriatric populations. The instruments also should ideally be usable and understandable by not only the therapist, but also the client, significant others in the client's life, insurance companies, and researchers. It should thus be clear and direct enough such that it can be understood by both a professional and a nonprofessional audience. Because it is often given over several different administrations, it should be sensitive to clinically important levels of change. As with any psychological test used by

clinicians, it should also have excellent psychometric properties. Finally, interpreting the results should be uncomplicated and the construct clear enough to enable feedback to the client or other relevant persons.

In addition to the many time-honored instruments such as the BDI, there have been more recent instruments designed specifically for treatment planning and patient tracking. For example, the Outcome Questionnaire (OQ-45; Lambert et al., 1996) is a 45-item, self-report instrument that requests clients to rate various areas on a five-point Likert scale. It can be used as an overall measure of client functioning, to establish a baseline, to assist with treatment decisions, and assess common symptoms (stress, *DSM-IV* V codes). The results are organized around level and type of symptom distress, interpersonal relations, and relative satisfaction with social role. The Butcher Treatment Planning Inventory (Butcher, 1998) is a 210-item self-report inventory that assesses issues and challenges that might be particularly relevant to treatment. The scales are organized around validity, treatment issues, and current symptoms. For example, the validity scales measure the extent to which the client has an overly virtuous presentation of himself or herself or tends to be closed-minded related to his or her difficulties. Treatment issues include areas such as somatization of difficulties, low expectations regarding treatment, and narcissism. A final example, which is discussed in more detail in Chapter 14, is the Systematic Treatment Selection model (Beutler, Clarkin, & Bongar, 2000), which includes a software package (Beutler & Williams, 1999) and a clinician rating form (STS Clinician Rating Form; D. Fisher, Beutler, & Williams, 1999). While the preceding instruments show considerable promise, they have not been used or tested as widely as many of the other instruments.

The three instruments selected for this chapter each fulfill the criteria required for treatment planning, monitoring, and outcome assessment. They are time efficient, directly relevant to treatment planning, can be used to evaluate outcome, effective as screening instruments, relevant for a wide range of target groups, sensitive to change, and the constructs and information they provide are sufficiently clear so that feedback is easy to give. Given these qualities, they are among the most extensively used brief instruments in both clinical practice and research.

SYMPTOM CHECKLIST-90-R (SCL-90-R) AND BRIEF SYMPTOM INVENTORY (BSI)

The Symptom Checklist 90-R (SCL-90-R; Derogatis, 1994) and its shortened version, the Brief Symptom Inventory (BSI; Derogatis, 1993), are ideally suited to quickly assess a client's type and severity of self-reported symptoms. It should not be regarded as a personality measurement, but is more an assessment of the current level of a variety of symptoms as experienced over a one-week interval. The SCL-90-R was derived from the earlier Hopkins Symptom Checklist (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974), which in turn had its origins in the much earlier Woodworth Personal Data Sheet (Woodworth, 1918). As the name suggests, the SCL-90-R consists of a series of 90 descriptions of symptoms that a client rates in terms of their severity (ranging from 0 = Not at all, to 4 = Extremely). A sixth-grade reading level is required, and it usually takes between 12 and 15 minutes to complete. The symptoms are scored around nine

different dimensions of symptoms (i.e., Somatization, Obsessive-Compulsive) as well as three global indexes (i.e., Global Severity Index). The BSI is a short form of the SCL-90-R composed of 53 of the SCL-90-R items, and it provides scores on the same symptom dimensions and global indexes.

Scores on the SCL-90-R are transformed onto a profile sheet displaying the nine symptom dimensions and three global indexes. Similar to the MMPI-2, each score has a mean of 50 and a standard deviation of 10. One of the unique features of the SCL-90-R is that these scores can be compared with and plotted based on the following four normative groups:

1. Norm A: *Psychiatric outpatients* ($N = 1{,}002$; 425 male, 577 females) approximately two-thirds of whom were White, and the entire sample was slightly skewed toward the lower end of the socioeconomic scale.
2. Norm B: *Nonpatients* ($N = 1{,}000$; 494 males, 480 female) representing a stratified random sample from a large U.S. eastern state.
3. Norm C: *Psychiatric inpatients* ($N = 313$; two-thirds female) of whom 55.7% were White, 43.6% Black with a mean age of 33.1.
4. Norm E: *Nonpatient adolescents* ($N = 806$; 60% females, 40% males) ages between 13 and 18 ($M = 15.6$) from two schools and composed primarily of middle class Whites.

If clinicians wish to make comparisons with a nonpatient group, they can use Norm B. In other situations, it might be advantageous to compare a person with either an outpatient or an inpatient reference group. These norms, combined with the wide diversity of validity studies, suggests the SCL-90-R can be used with a wide variety of respondents including medical patients, adolescents, community nonpatients, as well as inpatients and outpatients. It is also available in more than 26 languages, and computer scoring, administration, and interpretation programs are available.

Scoring and normative comparisons for the BSI follow similar procedures as for the SCL-90-R. The norms used on both scales are the same for psychiatric outpatients, psychiatric inpatients, and nonpatients. The BSI, however, has a larger adolescent normative base composed of 2,408 middle-class students (58% Whites, 30% Blacks, 12% other) between the ages of 13 and 19 ($M = 15.8$) from six different schools. There have also been additional norms developed and published separately for the elderly (Hale, Cochran, & Hedgepeth, 1984) and adolescent students (Canetti, Shalev, & Denour, 1994).

Reliability and Validity

The reliability of the SCL-90-R has consistently been good. The manual reports that internal consistency for the nine symptom dimensions based on psychiatric outpatients ranged from a low of .79 for Paranoid Ideation to a high of .90 for Depression. Internal consistency for "symptomatic volunteers" was slightly lower and ranged from a low of .77 for Psychoticism to .90 for Depression (in Derogatis & Savitz, 1999). Test-retest reliability over a one-week interval ranged from a low of .78 for hostility to a high of

.90 for Phobic Anxiety. Most coefficients were in the mid 80s. As expected, test reliability was slightly lower over a 10-week interval and ranged between .68 for Somatization to .83 for Paranoid Ideation (in Derogatis, 1994; Derogatis & Savitz, 1999).

Reliability for the BSI is similar although slightly lower than the SCL-90-R. Internal consistency ranged between .71 for Psychoticism and .85 for Depression (Derogatis, 1993). Similarly, internal consistency for bereaved parents ranged between .74 for Psychoticism to a quite high .97 on the Global Severity Index (L. C. Johnson, Murphy, & Dimond, 1996). Test-retest reliability over a two-week interval was a low of .68 for Somatization to .91 for Phobic Anxiety (Derogatis, 1993). One noteworthy feature was that the test-retest reliability was a quite high .91 for the Global Severity Index, indicating that it is a stable measure over time. This is particularly important given that the BSI (and SCL-90-R) GSI indexes are frequently used over repeated administrations to monitor treatment and evaluate its outcome.

Well over 1,000 studies have been done on the SCL-90-R investigating its validity. For example, both MMPI and General Health Questionnaire (GHQ) measures were found to converge with expected dimensions on the SCL-90-R and diverge with other expected measures (see Derogatis, 1994; Schmitz, Kruse, Heckrath, Alberti, & Tress, 1999). The SCL-90-R Depression dimension has been found to have a high correlation (.80) with the Beck Depression Inventory (Peveler & Fairburn, 1990) and to detect depression equally as effectively (Choquette, 1990). Expected SCL-90-R profiles have also been found for a variety of diagnostic groups including depression, anxiety, panic, sexual dysfunction, and substance abuse (see Derogatis, 1994; Derogatis & Savitz, 1999). However, other studies have questioned the divergent validity of the SCL-90-R dimensions, and authors have suggested that it be used as a general indicator of distress (Cyr, McKenna-Foley, & Peacock, 1985; Vassend & Skrondal, 1999). This controversy is strongly apparent in the findings related to factor structure. On one hand, Derogatis (1994) has reported that factor analytic research has, with the exception of the Psychoticism dimension, matched the various dimensions of the SCL-90-R (Derogatis, 1994). In contrast, other research has reported anywhere from one to six factors depending on the type of population that has been studied (Cyr et al., 1985; Hayes, 1997; Piersma, Boes, & Reaume, 1994; Vassend & Skrondal, 1999). For example, Vassend and Skrondal generally found a four-factor solution, but this varied depending on the gender and level of negative affect of the sample. They concluded that there was a "profound structural indeterminancy problem" (p. 685). Cyr et al. add that the factor structure becomes particularly uncertain when evaluated beyond the boundaries of neurotic outpatients. In contrast, Hayes found support for a six-factor solution with college students (but not for the nine dimensions listed on the SCL-90-R). The equivocal research on the factor structure of the SCL-90-R suggests that the nine dimensions be interpreted tentatively because most research has not supported their independence.

Although the factor structure of the SCL-90-R has been equivocal, research assessing the sensitivity and specificity for various disorders has been generally supportive. For example, the SCL-90-R detected relevant symptoms among bulimics with a sensitivity of 77% and specificity of 91% (Peveler & Fairburn, 1990). Similar levels of sensitivity (72%) and specificity (87%) for detecting psychological difficulties related to diabetes were also noted. High scores on the Hostility, Paranoid Ideation, Somatization, and Obsessive-Compulsive dimensions were able to detect the presence of Cluster

A (paranoid, schizoid, schizotypal) and Cluster B (antisocial, borderline, histrionic, narcissistic) with a quite high sensitivity of 89% and even higher specificity of 97% (Starcevic, Bogojevic, & Marinkovic, 2000). As would be expected given research on a one-factor solution, the SCL-90-R has been found to effectively detect the general level of distress a person is experiencing (Derogatis, 1993, 1994). The SCL-90-R has also been found to be responsive to clinically significant change (Schmitz & Hartkamp, 2000). In contrast to this favorable research on the diagnostic utility of the SCL-90-R, the Psychoticism dimension was not able to discriminate between psychotic and nonpsychotic patients (Stukenberg, Dura, & Kiecolt-Glaser, 1990). This, in combination with low internal consistency, indicates that the Psychoticism dimension seems to have the weakest psychometric properties of all the SCL-90-R scales.

Validity for the BSI is, in part, supported by the high correlations between the SCL-90-R and BSI dimensions, which range from a low of .92 for Psychoticism to a high of .98 for Paranoid Ideation (Derogatis & Savitz, 1999). This is sufficiently high so that research on the SCL-90-R not only supports the BSI, but also the two tests can be used as alternate forms for each other. Additional studies support the sensitivity of the BSI to distress and suggest that it can be used to track the outcomes of various interventions (Derogatis, 1993; Derogatis & Savitz, 1999). For example, screening of recently diagnosed cancer patients indicated that the BSI was sensitive to varying levels of distress based on ratings using outside criterion measures (Zabora, Smith-Wilson, Fetting, & Enterline, 1990). Similarly, elevated scores on the BSI have been found for bereaved parents (L. C. Johnson et al., 1996). Ratings by experienced clinicians of the level of distress experienced by clients have also been found to have moderate correlations with the expected dimensions on the BSI (Morlan & Tan, 1998). However, correlations between the BSI and client self-ratings of level of satisfaction with psychotherapy were not correlated (Pekarik & Wolff, 1996).

Interpretation

In many ways, "interpretating" the SCL-90-R and BSI is straightforward because the data is descriptive rather than interpretive. In other words, overall severity of a client's symptoms can be assessed through the degree of elevation on the Global Severity Index. Similarly, the severity by which a client is sensitive to the criticisms of others can be gauged based on the relative elevation of the Interpersonal Sensitivity dimension. However, clinicians may also wish to extend beyond these straightforward descriptions based on their clinical knowledge. For example, a person scoring high on Interpersonal Sensitivity is likely to exaggerate criticisms, ruminate over these criticisms, experience irrational thoughts, have low self-esteem, and be low in assertiveness. This can be further investigated by taking into account additional data. There may also be patterns of elevations that are consistent with various personality disorders. Avoidant personalities, for example, would be expected to have high scores on Interpersonal Sensitivity, Anxiety, and possibly Phobic Anxiety. In contrast, histrionic personalities would be likely to have elevations on Somatization. These conceptual links can be used as beginning points for further investigation to see if the person does or does not have the suggested personality styles. However, these "interpretations" should be considered more as hypotheses given the questionable independence of the nine SCL-90-R/BSI dimensions.

Interpretation can begin with the *global* indexes and then proceed to the *dimensional* and *symptom/item* level. Accordingly, the following information is a listing and elaboration on the meanings of the elevations in categories under these three general groupings (adapted from Derogatis & Savitz, 1999).

Global Indexes

Global Severity Index (GSI) This index is a combined rating that takes into account the intensity of experienced stress along with the number of reported symptoms. As such, it is the best single indicator of distress and should be used when a single measure is appropriate. A general rule of thumb is that a T-score above 63 suggests the presence of a clinically significant level of psychological difficulties.

Positive Symptom Distress Index (PSDI) The PSDI is an average rating for all symptoms that have been endorsed. Thus, it is a measure of symptom intensity (rather than merely the number of symptoms endorsed).

Positive Symptom Total (PST) Whereas the PSDI is a measure of symptom severity, PST represents the number (or breadth) of symptoms. Thus, a client could theoretically have a low PSDI, indicating that the symptoms they had were not particularly troubling, but might have a high PST indicating that they had a wide, potentially complex array of symptoms.

Symptom Dimensions

Somatization (SOM) An elevation on SOM indicates that distress is primarily experienced through concerns related to actual, amplified, or imagined physical dysfunction. Complaints might be focused on cardiovascular, gastrointestinal, respiratory, gross musculature, or other bodily areas (note responses to actual items). Pain and anxiety are both likely to be present as well, thereby amplifying any physiologically-based disorders. Interventions might involve increasing a client's awareness of how he or she uses somatization as a coping mechanism combined with alternative methods of coping such as stress management, social skills training, hypnosis, or biofeedback.

Obsessive-Compulsive (O-C) This dimension focuses on impulses, thoughts, and actions that are irresistible, repetitive, unwanted, and experienced as beyond the person's control. Some of the items also refer to more general cognitive performance deficits (i.e., the person's mind is going blank or he or she has trouble concentrating).

Interpersonal Sensitivity (I-S) High scores on I-S indicate the person has considerable discomfort in interpersonal situations. They have negative expectations regarding relationships and are self-conscious. When they compare themselves with others, they typically feel inferior and, thus, experience self-doubt and inadequacy. Crucial to any intervention is a supportive therapeutic relationship combined with cognitive restructuring and assertiveness training.

Depression (DEP) Elevations on DEP indicate the person is experiencing a range of depressive symptoms. These might include loss of pleasure, dysphoria, loneliness, crying, withdrawal, pessimism, sleep disturbance, alterations in appetite, poor motivation,

and low energy (check individual items). There may also be the presence of suicidal ideation and other cognitions consistent with depression.

Anxiety (ANX) This dimension focuses on the presence of apprehension, nervousness, trembling, and dread. This may or may not be consistent with panic attacks. Physiological components of anxiety, including rapid heart rate, tension, and restlessness, are also likely to be present. Possible interventions include relaxation training, stress management, assertiveness training (and other forms of skills training), and exercise.

Hostility (HOS) Persons scoring high on HOS experience resentment, irritability, aggression, and, possibly, rage. Accordingly, anger management might be an appropriate recommendation.

Phobic Anxiety (PHOB) This dimension focuses on the presence of excessive and irrational fear related to a person, place, object, or situation. He or she might report a fear of open places, anxiety when traveling away from familiar areas, and fear of developing a panic attack. Although the title of the dimension appears to be related to phobias, most of the actual items reflect the more pathological aspect of phobias to the extent that high scores may reflect agoraphobia or panic attacks rather than merely phobias. Interventions can be focused on the areas of greatest anxiety and might include graded exposure, relaxation training, hypnosis, and cognitive restructuring.

Paranoid Ideation (PAR) Items in this dimension tap into the key dimensions of paranoid thought including hostility, projection, grandiosity, suspiciousness, and a need for control based on a fear of losing independence. Delusions may also be present and are reflected in items related to fears of being watched, talked about, or not being given credit for achievements.

Psychoticism (PSY) High scores reflect a person who is extremely withdrawn and isolated and may be experiencing core symptoms of schizophrenia, including hallucinations (hearing voices, thought broadcasting) and thought control. Scores can be seen as being on a psychoticism continuum ranging from minor levels of interpersonal alienation to a full display of severe psychotic symptoms.

T scores above 63 on two or more of the preceding dimensions suggest that the person has clinically significant levels of psychological distress. Additional potentially important symptom-related items not scored on these dimensions include those related to poor appetite, sleep disturbance, fear of dying, overeating, early morning awakening, difficulties with sleep maintenance, and guilt. These can be noted to obtain additional information. Researchers have also developed additional scales that may be used some time in the future to extend interpretation (see the SCL-90-R Mania Scale; E. Hunter et al., 2000 and factor-based scales for college students; Hayes, 1997).

Symptom Level/Item

Additional information can be obtained by noting the content of the individual items the client has endorsed. For example, items on the Depression dimension can provide

specific information related to the person's depression. Importantly, the presence of Suicidal Ideation (i.e., item related to ending the person's own life) can be noted and should then be followed up by more in-depth assessment for risk of self-harm. The presence and extent of possible vegetative symptoms (low energy, sleep problems, loss of sexual energy) can also be noted; this may have implications for various treatment recommendations. Items that the client has answered either "quite a bit" or "extremely" can be considered critical. These should be given particular attention for assessment, treatment planning, and establishing a relevant baseline and outcome to treatment.

THE BECK DEPRESSION INVENTORY

The Beck Depression Inventory (BDI) was introduced in 1961 by A. T. Beck, Ward, Mendelson, Mock, and Erbaugh, was revised in 1971, and was copyrighted in 1978 (A. T. Beck, Rush, Shaw, & Emery, 1979). Although the later version, referred to as the BDI-IA, involved a clarification and modification of the items, the two versions were found to be highly correlated (.94; Lightfoot & Oliver, 1985). The BDI underwent a further and major revision in 1996 (BDI-II) to include a wider range of symptoms (A. T. Beck et al., 1996). By so doing, it became more congruent with *DSM-IV* diagnostic criteria for depressive disorders. Four of the items were replaced to reflect symptoms consistent with more severe depression (Agitation, Worthlessness, Concentration Difficulty, and Loss of Energy). A further two items were revised to better reflect decreases in appetite and sleep. In addition, many of the other items were reworded.

Comparisons between the BDI/BDI-IA and the BDI-II indicate that clients are likely to endorse one to two more items/symptoms on the BDI-II when compared with the earlier BDI and BDI-IA (A. T. Beck et al., 1996; Dozois, Dobson, & Ahnberg, 1998; Steer, Rissmiller, & Beck, 2000). More symptoms are likely to be endorsed toward the higher ranges of depression (three or more items/symptoms) than the lower ranges. Using an outpatient sample, BDI-IA/BDI-II correlations were .84, and the mean total scores were slightly higher for the BDI-II than the BDI-IA (21.63 versus 18.15; A. T. Beck et al., 1996). Correlations between the BDI and BDI-II for a university population indicated a slightly higher correlation of .92 (Dozois et al., 1998). Despite the slightly higher scores on the BDI-II, this information indicates that the BDI-II is sufficiently comparable to its predecessors such that, with appropriate caution, much of the research on the BDI/BDI-IA can be generalized to the more recent BDI-II.

The BDI-II and its predecessors have been widely used for the assessment of depression among psychiatric patients (Camara et al., 2000; C. Piotrowski, 1996; Steer, Ball, Ranieri, & Beck, 1999; Steer et al., 2000) as well as depression in normals (A. T. Beck et al., 1996; Steer, Beck, & Garrison, 1986). It has been found to detect depression as effectively as longer and more costly structured interviews (Stukenberg, Dura, & Kiecolt-Glaser, 1990). The popularity of this instrument is amply demonstrated in that, in the 40 years since its introduction, well over 1,000 research studies have been performed either on or using it.

The items in the BDI were originally derived from observing and summarizing the typical attitudes and symptoms presented by depressed psychiatric patients (A. T. Beck et al., 1961). A total of 21 items related to various symptoms were included,

and, when completing the inventory, respondents are requested to rate the intensity of these symptoms on a scale from 0 to 3. Typical questions relate to areas such as sense of failure, guilt feelings, irritability, sleep disturbance, and loss of appetite. The inventory is self-administered and takes from 5 to 10 minutes to complete. A fifth- to sixth-grade reading level is required to adequately comprehend the items. The total possible range of scores extends from a low of 0 to a theoretical high of 63. However, only the most severe levels of depression are reflected by scores of 40 or 50. More typically, clinically depressed or maladaptively nonclinical populations score in the 14 to 28 range (A. T. Beck et al., 1996).

Reliability and Validity

Since its initial development in 1961, the BDI has been subjected to extensive psychometric evaluation. A meta-analysis of the original BDI/BDI-IA indicated that internal consistency ranged from .73 to .92 with a mean of .86 (A. T. Beck, Steer, & Garbin, 1988). Test-retest reliabilities have ranged from .48 to .86, depending on the interval between retesting and type of population (Beck et al., 1988). However, repeat administrations over seven weeks at one administration per week using university students indicated a 40% decline in scores (Ahave, Iannone, Grebstein, & Schirling, 1998). This suggests that some of the reduction in scores for clinical populations following interventions may be partially accounted for (approximately 10% of the variance) by a natural reduction rather than the intervention itself. Research with the BDI-II has consistently found a high internal consistency ranging from .89 to .94 even when using a variety of populations (Arnau, Meagher, Norris, & Bramson, 2001; A. T. Beck et al., 1996; Dozois et al., 1998; Steer et al., 1999, 2000). Test-retest reliability over a one-week interval was .93 (Beck et al., 1996).

Evaluation of content, concurrent, and discriminant validity as well as factor analysis has generally been favorable. The content of the BDI items was derived by consensus from clinicians regarding symptoms of depressed patients combined with considerations related to the various *DSM-IV* categories for the diagnosis of depression. Concurrent validity is suggested by high to moderate correlations with clinical ratings for psychiatric patients (A. T. Beck et al., 1996). In addition, moderate correlations have been found with similar scales that also rate depression, such as the Hamilton Psychiatric Rating Scale for Depression (.71), Beck Hopelessness Scale (.68; Beck et al., 1996), and the Depression Anxiety Stress Scale (.88; Osman et al., 1997). The BDI has been able to discriminate psychiatric from nonpsychiatric populations (Beck et al., 1996) as well as discriminate the level of adjustment in psychiatric populations (Arnau et al., 2001; Beck et al., 1996). The BDI-II's ability to discriminate between primarily anxiety as opposed to primarily depressive disorders is supported in that BDI-II scores were more highly correlated with the Hamilton Psychiatric Rating Scale for Depression (.71) compared with the Hamilton Rating Scale for Anxiety (.47). Similarly, Steer et al. (2000) found higher correlations between the BDI-II and the SCL-90-R Depression dimension (.89) than the SCL-90-R Anxiety dimension (.71).

Factor analytic studies indicate that the BDI is composed of a Noncognitive (or Somatic-Vegetative) factor comprising contents related to somatic aspects of depression (loss of energy, changes in sleep patterns, crying) and a Cognitive-Affective factor

related to self-reported thoughts clients use to describe their attitudes toward themselves and their depression (self-dislike, suicidal thoughts, thoughts of worthlessness; A. T. Beck et al., 1996). These factors have been found to be consistent among various samples including college students (Beck et al., 1996), adolescents (Steer, Kumar, Ranieri, & Beck, 1998), geriatrics (Steer et al., 2000), primary care medical patients (Arnau et al., 2001), geriatric inpatients (Steer et al., 2000), and clinically depressed outpatients (Steer et al., 1999). Osman et al. (1997) found a slightly different factor structure composed of Negative Attitude, Performance Difficulty, and Somatic Elements using a sample of undergraduates. Most comparisons between the BDI and BDI-II indicate that the factor structure on the BDI-II is more clearly defined, suggesting it is a slightly superior instrument (Dozois et al., 1998).

Interpretation

The following scores can be used to indicate the general level of depression:

0 to 13	No or minimal depression
14 to 19	Mild
20 to 28	Moderate
29 to 63	Severe
Below 4	Possible denial of depression, faking good; lower than usual scores even for normals

Scores significantly above even severely depressed persons suggest possible exaggeration of depression; possibly characteristic of histrionic or borderline personality disorders. Significant levels of depression are still possible. Arnau et al. (2001) found that a cutoff score of 18 correctly classified 92% of patients with major depressive disorder.

An ipsative interpretation of BDI responses can be used to specify irrational beliefs and relevant symptoms that are likely to be related to a person's depression. Identification of these beliefs and symptoms can be useful in specifying areas that need to be worked on in therapy. Any of the following (A. T. Beck et al., 1996, p. 5) can be assumed an area of difficulty if a score of 3 is indicated on the numbered item:

1. Sadness.
2. Pessimism.
3. Past failure.
4. Loss of pleasure.
5. Guilty feelings.
6. Punishment feelings.
7. Self-dislike.
8. Self-criticalness.
9. Suicidal thoughts or wishes.
10. Crying.
11. Agitation.

12. Loss of interest.
13. Indecisiveness.
14. Worthlessness.
15. Loss of energy.
16. Changes in sleeping pattern.
17. Irritability.
18. Changes in appetite.
19. Concentration difficulty.
20. Tiredness or fatigue.
21. Loss of interest in sex.

One specific area to be alerted to is the potential for suicide, which can be indicated by strong endorsements (2 or 3) on items 9 (suicidal thoughts or wishes) and 2 (pessimism). Whereas the level of depression (based on total score) and presence of specific item endorsement can assist in suggesting the presence of a formal *DSM-IV* disorder, a definitive diagnosis would still need to be made based on a more thorough review by a clinician.

STATE TRAIT ANXIETY INVENTORY (STAI)

A client's level of anxiety is one of the most crucial dimensions to assess in treatment planning as well as to establish the impact of interventions. The State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is ideally suited for this purpose because it is a brief (20 item), self-report inventory that is easy to understand, and is sensitive to transitory episodes of anxiety (states) as well as more stable personality features that predispose a client to experiencing more chronic levels of anxiety (traits). It is currently the most frequently used measure of anxiety with well over 8,000 studies available in the literature. Research has evaluated its use in the treatment of phobias, test anxiety, panic, generalized anxiety, and the impact of specific types of treatment such as cognitive behavior therapy, systematic desensitization, relaxation, and rational emotive therapy (Spielberger, Sydeman, Owen, & Marsh, 1999). It has also been used extensively in cross-cultural research and has been translated into more than 58 languages and dialects.

Construction of the STAI began in 1964 with a single set of items that could be used to assess either state or trait anxiety based on rewording the instructions (Form A). The state instructions requested the client to complete items for how they felt "right now, at the moment" whereas the trait descriptions asked them to indicate how they generally feel. The items were originally derived and adapted from existing anxiety inventories including the Affect Adjective Checklist (see Spielberger et al., 1999). Items were reduced and the scale refined based on the degree to which individual items correlated with the Manifest Anxiety Scale, Anxiety Scale Questionnaire, and Welsh Anxiety Scale of the MMPI (see Spielberger et al., 1999). Further evaluation with Form A indicated that merely rewording the instructions was not sufficient to eliminate the clear

trait connotations of some of the items. For example, the item "I worry too much" was a good measure of trait anxiety but merely rewording the instructions indicated that it was not a good measure of state anxiety. As a result, a second form (Form X; see Spielberger et al., 1999) was developed based on the trait and state dimensions having their own individual items. Trait items were selected based on their having the highest correlations with the Manifest Anxiety Scale, Anxiety Scale Questionnaire, and Welsh Anxiety Scale as well as being the most stable over time. The state items were selected based on their being most sensitive to high versus low stress conditions (high construct validity) and having the highest internal consistency.

A decade after the publication of Form X, the STAI underwent a further major revision based on factor analysis, a clearer understanding of the concept of anxiety, and an attempt to eliminate item overlap with depression. This resulted in the current (Form Y) version having 10 items for trait and 10 for state anxiety (Spielberger et al., 1983). Form Y was normed on 1,838 employees of the Federal Aviation Administration, 855 university students, 424 high school students, 1,701 Air Force recruits, and 263 naval recruits. Older persons and those with more education scored somewhat lower than those who were either younger or less educated, which suggests it might be important to use age and education-related norms. Additional norms are available for a neuropsychiatric population, general medical/surgical patients, and young prison inmates. A children's form, the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973), is also available.

Reliability and Validity

Test-retest reliability for college students over 30- and 60-day intervals indicated reasonably good coefficients ranging between .73 and .86 for trait anxiety. In contrast, state anxiety test-retest reliabilities were relatively lower, ranging from .51 for males to .36 for females (Spielberger et al., 1983). The lower range for state anxiety is expected given that state anxiety is considered a more changeable construct. Given the expected fluctuations for state anxiety, measures of internal consistency would be more appropriate and important to consider. These have resulted in quite high state anxiety median coefficients ranging between .88 and .93 with a similarly high median trait anxiety coefficient ranging from .92 to .94 (Kabacoff, Segal, Hersen, & Van Hasselt, 1997; Spielberger et al., 1983).

The content validity of the STAI-Trait scale is supported in that five out of a possible eight domains for a *DSM-IV*-based diagnosis of a generalized anxiety disorder were reflected in the items (Okun, Stein, Bauman, & Silver, 1996). Concurrent validity is supported in that correlations with the Manifest Anxiety Scale and the Anxiety Scale Questionnaire have ranged from .73 to .75 (Spielberger et al., 1999). These correlations are sufficiently high that the STAI can be considered alternative measures of trait anxiety. However, the STAI has the advantage of being shorter and less contaminated by measures of depression. Lower and moderate, but still significant, correlations were found between the STAI-Trait and the Worry Scale (.57) and Padua Inventory (.57; Stanley, Beck, & Zebb, 1996).

The construct validity of the STAI is suggested in that psychiatric patients generally have higher scores on trait anxiety than nonpatient groups (Spielberger et al., 1983;

Stanley et al., 1996). One exception is that, as expected, patients with character disorders tended to have lower scores (Spielberger et al., 1983). Kabacoff et al. (1997) also found that patients with anxiety disorders had slightly higher STAI-Trait scores than patients without anxiety disorders. Despite this support for the convergent and divergent validity of the STAI, Kabacoff et al. were not successful in developing adequate cutoff scores for identifying the presence of an anxiety disorder. This was primarily because of difficulty finding a score that produced both good sensitivity (high identification of true positives) as well as good specificity (high identification of true negatives). Construct validity for the validity of STAI-State anxiety is supported in that students during in-class exams and military recruits undergoing stressful training procedures had higher scores when compared to scores taken after relaxation procedures or with age-matched controls (Spielberger et al., 1983). Numerous studies have demonstrated that the STAI is sensitive to the impact of a wide variety of interventions (Spielberger et al., 1983; Spielberger et al., 1999).

Factor analysis of the STAI has been mixed. According to the STAI scale development, there should ideally be one factor that loads on trait anxiety and another one on state anxiety (Spielberger et al., 1983). In contrast, Bieling, Antony, and Swinson (1998) found a higher order factor derived from the trait anxiety items they referred to as *negative affect* and two lower order factors that they concluded were organized around *depression* and *anxiety.* Thus, the trait items seemed not to be pure measures of anxiety but included measures of negative affect and depression, as well as anxiety. Whereas Spielberger et al. (1983) did attempt to make Form Y more of a pure measure of anxiety than Form X, this seems to have been only partially successful. This underlies the issue, frequently found in other measures of anxiety and depression, that anxiety and depression have overlapping features with correlations typically ranging between .45 and .75 (Lovibund & Lovibund, 1995). The factor structure of the STAI is further complicated in that Kabacoff et al. (1997) found two factors related to whether the items were worded in a positive or negative direction. They concluded that these items were "method factors" unrelated to the constructs of anxiety.

Interpretation

Because the STAI comprises two unidimensional subscales, interpretation is mainly composed of descriptions for the variables being measured. This can be done by considering both the variable itself and the relative magnitude of the person's score.

High T-Anxiety The person is likely to perceive a wide number of situations as threatening or dangerous; they are especially likely to be concerned with being evaluated by other people with corresponding threats to their self-esteem.

High S-Anxiety The person has feelings of apprehension, worry, nervousness; unpleasant, consciously perceived feelings of tension; the person is also likely to report corresponding activation of the autonomic nervous system.

High S-Anxiety/Low T-Anxiety The anxiety the person is reporting is likely to be caused by some external threat or a current situational stressor. As a result, it is likely to

resolve itself. If intervention is warranted, it should ideally be directed toward strategies that provide a reduction in arousal, such as increasing their social supports, systematic desensitization, providing reassurance, hypnosis, exercise, meditation, or progressive muscle relaxation. There might also be an emphasis on what the current anxiety has taught them about themselves and how this might be used to reduce the likelihood of reducing anxiety in the future.

High T-Anxiety/Low S-Anxiety Although the person is not currently reporting anxiety, he or she is prone to reacting to situations in such a way as to easily become anxious. He or she is likely to be extremely concerned with threats to self-esteem and, as a result, might be apprehensive in any interpersonal situation in which he or she might be judged.

RECOMMENDED READING

Antony, M. M., & Barlow, D. (Eds.). (2002). *Handbook of assessment, treatment planning, and outcome evaluation: Empirically supported strategies for psychological disorders.* New York: Guilford Press.

Maruish, M. E. (Ed.). (1999). *The use of psychological testing for treatment planning and outcome assessment* (2nd ed.). Hillsdale, NJ: Erlbaum.

Maruish, M. E. (2002). *Psychological testing in the age of managed behavioral health care.* Mahwah, NJ: Erlbaum.

Stout, C. (1997). *Psychological assessment in managed care.* New York: Wiley.

Chapter 14

PSYCHOLOGICAL ASSESSMENT AND TREATMENT PLANNING

The ultimate goal of psychological assessment is to help solve problems by providing information and recommendations relevant to making the optimum decisions related to the client. This involves integrating a wide variety of information including specifics of the problem, client resources, a client's personal characteristics, and environmental circumstances. Practitioners must then work with this information to make recommendations related to treatment setting (inpatient/outpatient), intensity (frequency and duration), goals, mode (individual, group, family), and specific strategies and techniques. The sheer number of these variables can make assessment a daunting task. Thus, the focus of this chapter is to provide a manageable framework for systematically organizing assessment results for planning treatment. Research has shown that following this framework optimizes the outcome of interventions.

The following format for organizing results and developing treatment plans has been guided by several principles and values. When possible, empirically supported information has been provided. This is possible using the knowledge derived from the rather extensive body of research currently available. In fact, treatment that ignores the procedures indicated by current research runs the risk of not offering clients the most effective treatments available. At the same time, it is acknowledged that clinical experience and judgment inevitably needs to interact with the research, assessment results, and the uniqueness of the client to generate the best treatment plan. A further guiding principle underlying this chapter is that the format is both sequential and systematic. It is sequential in that, typically, a series of decisions confront clinicians beginning with areas such as how restrictive interventions should be and ending with issues such as specific techniques of therapy and methods of relapse prevention. Finally, the number of variables considered has been reduced to those that seem most relevant, easily manageable, and best supported by research.

Developing effective recommendations requires a number of knowledge and skill areas beyond merely test interpretation. One of the more important ones relates to general case management. This requires practitioners to survey the general case issues, focus on the most salient features, and make recommendations accordingly. This should include noting how restrictive treatment should be, which is directly related to the severity of the problem and whether the patient is likely to present a danger to self or others. After reviewing these considerations, practitioners need to be aware of the resources available in the community and make recommendations to the

most appropriate one(s). This might include treatment in a specific inpatient setting or referral to areas such as an outpatient clinic, medical facility, suicide prevention center, Alcoholics Anonymous, or a behavioral medicine unit. Decisions also need to be made related to the frequency and duration of treatment. Practitioners should also be able to assess and provide recommendations on how to optimize a client's environment. For example, assessing the client's level of social support might help either in encouraging the person to use available supports, or in enhancing only partially adequate supports. Environments might also be changed to increase social interaction or decrease the likelihood of relapse.

Practitioners can and should be able to deliberately tailor their responses toward specific characteristics and circumstances of the client. While this might seem self-evident, many therapists typically provide the same or at least similar interventions for all their clients. Frequently, these interventions are based on the specific school of therapy the therapist is most familiar with (e.g., cognitive therapy for every client who comes in for treatment). Research has demonstrated, however, that whereas cognitive behavior therapy can be effective for patients with externalizing coping styles, a supportive, self-directive method is more effective for patients with internalizing styles of coping (Beutler et al., 2000). A further assumption frequently found in clinical lore is that empathy is an essential ingredient of all effective therapy. Despite this, controlled studies indicate suspicious clients with low motivation do poorly when psychotherapists are empathic, involved, and accepting (Beutler, Harwood, Alimohamed, & Malik, in press; Beutler, Crago, & Arizmendi, 1986). These examples, and many others, indicate that therapists need to have relational flexibility and a broad range of skills. In contrast, providing clients with a narrow range of possible interventions not only may reduce treatment effectiveness, but also may raise a question of ethics because the best interventions are not being provided for them. Recommendations and interventions should, as much as possible, be guided by research as clinical lore can sometimes be misleading.

This brief introduction to treatment planning is not intended to minimize either the tremendous impact that the quality of the treatment relationship has on outcome, or of the importance of clinical experience. The overall quality of the therapeutic relationship accounts for at least as much of the outcome variance as specific techniques (Ahn & Wampold, 2001; Blatt, Zuroff, Quinlan, & Pilkonis, 1996; Horvath & Symonds, 1991; Wampold, 2000). Well-defined techniques, however, are often easier to specify and control than the more general quality of the relationship. In addition, techniques that match a client's needs and expectations are likely to enhance the quality of the working relationship. Thus, it is difficult, if not impossible, to separate technique and relationship. For example, relationship quality is likely to deteriorate if a therapist tries highly directive techniques with quite defensive clients (Beutler, Moleiro, & Talebi, 2002; Beutler, Sandowicz, Fisher, & Albanese, 1996). In addition, clinical experience will always be crucial in integrating a diverse range of client information into an optimum set of recommendations. While this process should be generally guided by available research, the specifics of a particular case might be sufficient to alter or even negate the generalities suggested by research data alone. Thus, research findings and clinical information should ideally be in an active interplay such that they optimize each other's strengths and minimize their respective weaknesses.

DEVELOPMENT AND APPROACHES TO TREATMENT PLANNING

One of the central concerns for researchers and clinicians refining treatment planning has been efforts to understand how and why therapeutic interventions do or don't work. Similar to the debates on intelligence, researchers and clinicians can be divided into "splitters," who have focused on the impacts of specific techniques, or "lumpers," who have been more concerned with the common, nonspecific ingredients that facilitate change. A further related theme is the identification of relevant client domains or behaviors needing change and matching these with appropriate interventions. The general purpose of assessment in this process is to identify the most relevant client characteristics or symptom behaviors and match these with optimal interventions. Gordon Paul (1967) ambitiously stated this agenda with a question: "*What* treatment, by *whom,* is most effective for *this* individual with *that* specific problem, and under *which* set of circumstances?" (p. 44).

Ancient traditions of mental health were fully aware of the importance of tailoring interventions toward the specifics of the client. For example, the Vedas discuss the differential effects of telling appropriate metaphors to clients according to their needs. Similarly, Sufism has had a well-developed tradition of storytelling designed to create specific impacts on the participants (Groth-Marnat, 1992). As early as 1919, Freud was concerned with matching patients to different types of psychotherapy. Classical psychoanalysis was recommended for patients who were quite psychologically minded. In contrast, clients who were considered "unanalyzable" because of a lack of psychological sophistication were referred for psychoanalytic psychotherapy that focused on direct suggestion rather than extensive insight and in-depth self-exploration.

Throughout the 1950s and 1960s, an extremely diverse number of therapies were developed. Each one provided a different theoretical model for causation and a wide variety of techniques. Part of what stimulated these developments was the hope that a series of techniques would prove successful in treating certain types of problems. Examples of such techniques included systematic desensitization for phobias or interpreting the transference as a tool in resolving past interpersonal conflicts. In the psychosomatic literature, it was believed that certain disorders (e.g., asthma) were the result of specific types of conflicts (e.g., suppressed dependency needs). Resolving these specific conflicts, it was hoped, would similarly remove the relevant symptoms. This extensive variety and specificity has led to the development of more than 400 different types of psychotherapies, only a few of which have been subjected to any degree of empirical investigation.

Psychological assessment during the 1950s and 1960s closely paralleled the particular school of therapy it was aligned with. Because many assessment procedures were both used in a medical context and relied-on projective techniques, they, accordingly, reflected a psychoanalytic perspective. The goal, then, was to list a patient's symptoms along with a dynamic interpretation of the conflicts believed to be causing these symptoms. The specificity of treatment planning was de-emphasized in favor of detailed descriptions of inner dynamics. It was assumed that, by describing these conflicts, the therapist would then know better how to proceed. During the 1960s and 1970s, the competing schools of behaviorism and humanism developed their own modes of assessment based on either specifying target behaviors and the antecedent events leading to these

behaviors, or attending to the ongoing experience of the client. In either case, the value of traditional psychometric procedures was not only de-emphasized, but even criticized and abandoned.

Understandably, there was considerable competition between the different therapies as to which one was most effective. In 1952, Eysenck stimulated considerable controversy with his verdict that psychotherapy (particularly psychoanalysis) was no more effective than placebo. In contrast, he concluded that behavior therapy has demonstrated positive outcomes beyond merely placebo effects (Eysenck, 1994). Much of the ensuing research became a horse race in which proponents of particular schools wanted to demonstrate the superiority of the chosen therapeutic mode that they had received training in for so many years. The classic and much-cited summary study of therapeutic outcome was M. L. Smith, Glass, and Miller's (1980) meta-analysis, which concluded that all of the evaluated therapies were effective. They also found greater effect sizes for those therapies with a progressively narrow focus than for those with a wider focus. For example, techniques such as systematic desensitization and hypnosis, which typically target a narrow band of behavior (elimination of a phobia, habit modification), were found to have greater impact than client-centered therapy, with its more general goal of personal growth. However, the differences between the various therapies were not extensive, which led many reviewers of the field to agree with Luborsky, Singer, and Luborsky's (1975) earlier verdict that "Everybody has won and all must have prizes" (often referred to as the "dodo bird" verdict). This is supported by more recent, methodically well-designed studies that have demonstrated little or no differential outcomes between different therapies when targeted at the same problems (Ahn & Wampold, 2001; E. Anderson & Lambert, 1995; Patterson, 1989; Seligman, 1995; Wampold, 2000). For example, current high-quality research (including a meta-analysis) has found that randomly assigned manualized cognitive-behavioral versus psychodynamic-interpersonal interventions for depression had similar effectiveness on therapeutic outcome (Gallagher-Thompson & Steffen, 1994; Leichsenring, 2001; D. A. Shapiro et al., 1994).

The preceding studies, along with responses to these findings, have significant implications for treatment planning. One category of response is an investigation of the *nonspecific features of therapy* common to all systems (see Ahn & Wampold, 2001; Andrews, 2001). Underlying this response is the hope that these nonspecific factors would explain the general equivalence of outcomes across therapies. The earliest formal conceptualization was a 1957 description of "necessary and sufficient conditions of therapeutic change" by C. Rogers (1957/1992). These included genuineness, unconditional positive regard, and accurate empathy. A somewhat similar nonspecific formulation was also proposed by J. Frank (1973), who emphasized that successful therapy involved providing the client with hope, overcoming demoralization, and creating a corrective emotional experience involving benevolent persuasion. This nonspecific focus provides a contrast to the more directive, technique-oriented approaches. In particular, the nonspecific explanations place considerable emphasis on the quality of the therapeutic relationship beyond mere technique (see Norcross, 2002). The implications for assessment and treatment planning are that the technical aspect of assessment (formal tests) recedes in importance compared with the quality of the relationship (Andrews, 2001; Luborsky, 1994). Formal testing may even be perceived as interfering with the development of a positive therapeutic relationship. In addition, the specificity of

treatment recommendations is also de-emphasized. What still remains, however, are basic case management issues (restrictiveness, format, and intensity of treatment) and enhancing aspects of the relationship that are likely to maximize outcome (i.e., matching client expectations, being perceived as trustworthy and credible).

A second general strategy has collectively been referred to as *differential therapeutics.* This approach focuses on refining intervention techniques based on specific diagnoses combined with additional information related to aspects of the problem (see Antony & Barlow, 2002; Nathan & Gorman, 1998; Sammons & Schmidt, 2001). The general function of assessment in differential therapeutics is to diagnose and evaluate the specifics of a disorder as carefully as possible. Techniques believed to be most effective in optimizing outcome are tailored and directed toward a symptom or symptom cluster. This model closely parallels and draws on procedures used in medicine, which similarly rely on accurate diagnosis before applying the optimal treatment.

The preceding approach has had varying degrees of success. Probably the most noteworthy of these successes has been the development of specific targeted interventions for clusters of anxiety-related symptoms (Barlow, 1988; J. G. Beck & Zebb, 1994; Steketee, 1994). In particular, Barlow, Craske, Cerny, and Klosko (1989) have developed a specific targeted treatment for panic disorder that has been found to be effective for 80% to 100% of those who completed the program. In addition, it has been found to provide outcomes clearly superior to pharmacotherapy (Gould, Otto, & Pollack, 1995). The treatment involves a combination of muscle relaxation, cognitive restructuring, and exposure to internal sensations linked to training in breathing. Interventions for social phobia and social anxiety have also shown differential effectiveness over other forms of treatment. Such programs involve restructuring cognitions, simulations of feared situations, and homework assignments in which clients gradually expose themselves to actual anxiety-related situations (Hope & Heimberg, 1993). Finally, differentially effective interventions for obsessive-compulsive disorder have primarily centered on gradual exposure to the anxiety-related situations, along with strategies to prevent the occurrence of the compulsive behaviors (Riggs & Foa, 1993).

While most of the anxiety disorders have indicated the advantage of using interventions targeted directly at the subtype of disorder (diagnosis), less success has been achieved for specific interventions in the treatment of depression. The extent of vegetative symptoms, presence of manic episodes (bipolar), and presence of suicidal risk have implications for type of medication and restrictiveness of treatment. However, research has so far not been able to clearly identify the best psychosocial intervention for depression (Gallagher-Thompson & Steffen, 1994; Leichsenring, 2001; Rude, 1986; D. A. Shapiro et al., 1994) although some have argued for the differential effectiveness of cognitive behavioral approaches (see Antonuccio, Danton, & DeNelsky, 1995). Researchers have also had difficulty demonstrating differential effectiveness for specific psychosocial interventions for schizophrenia, sleep disorders, sexual disturbances, generalized anxiety disorder, and personality disorders (Beutler & Crago, 1986; T. Brown, O'Leary, & Barlow, 1993).

A third general response has been to consider the nonequivalence of therapeutic outcomes to be the result of insufficiently explored *client characteristics* (see Beutler, 1979; Beutler et al., 2000; Beutler & Harwood, 2000). This would mean that some types of clients do quite well when provided with a certain type of therapy and others,

given the same therapy, do quite poorly. If those clients who did poorly could have been identified and provided with different strategies, they might have made significant therapeutic gains using an alternate approach. However, the averaged scores on outcome studies using heterogeneous populations have obscured these potentially relevant client differences. The strategy, then, has been to thoroughly research a wide variety of client characteristics to determine which ones can be used to predict differential response to therapy. Over 200 of these characteristics have been suggested, of which 100 have been subjected to empirical investigation (Beutler et al., 2000; Garfield, 1994; Norcross, 1997). The result has been that, over the past 20 years, there has been increasing delineation and use of the most empirically validated characteristics for systematic treatment planning (Beutler et al. 2000; Beutler & Clarkin, 1990; Harwood & Williams, 2003). Reviews of this strategy have indicated that, under optimal matching conditions, up to 64% of the outcome variance can be accounted for (Beutler, 1983, 1989; Berzins, 1977). When client characteristics and treatment matching are combined with the quality of the therapeutic alliance, prediction of outcome increases to 90% (Beutler et al., 1999; Beutler, Moleiro, Malik, & Harwood, in press). In contrast, providing therapeutic techniques without considering predisposing client characteristics has been found to account for only 10% of the outcome variance (Beutler, 1989; Wampold, 2000). The implication for assessment is that predisposing client characteristics can and should be used to identify relevant dimensions. Furthermore, these dimensions should then be used to develop optimum treatment plans. This emphasizes both the technical and clinical aspects of assessment as well as the specificity of treatment recommendations. This does not negate the importance of common factors (caring, empathy, respect, etc.), but systematic treatment selection can potentially add to the effects of these common factors.

In addition to the preceding three general strategies, a variety of specific attempts has emerged to provide guidelines for prescriptive matching of client characteristics with therapeutic interventions. Ideally, the *DSM* should be useful in developing treatment plans in a similar manner as occurs for specific disease entities in general medicine. Generally, however, this has not been the case. Although some of the diagnostic categories have implications for different forms of somatic interventions (i.e., antidepressants for depressive disorders), they generally are not particularly helpful for designing psychosocial interventions (Beutler, 1989; Houts, 2002). In an effort to more clearly identify the full array of relevant domains for intervention, Lazarus (1973) suggested that clinicians analyze a patient's **B**ehaviors, **A**ffects, **S**ensory experiences, **I**magery, **C**ognitions, **I**nterpersonal relationships, and need for **D**rugs (BASIC-ID; see Chapter 4). A somewhat different perspective has been taken by authors who believe that the various stages of therapy or change are crucial to consider in tailoring interventions. Prochaska and DiClemente (1984, 1992) encouraged practitioners to tailor their interventions around the stages of precontemplation, contemplation, preparation, action, and maintenance.

In a behavioral medicine context, Wickramasekera (1995a, 1995b) has developed a high-risk model for identifying and assessing clients likely to have somatizing complaints. This includes predisposing factors consisting of either very high or very low hypnotizability, neuroticism (level of sympathetic reactivity), and catastrophizing cognitions. Precipitating factors relate to major life changes or minor hassles, and

client factors that are likely to serve as buffers include level of social support and coping ability. Treatment can then be tailored toward the patterns of scores on these client dimensions.

A further strategy has been to determine the factors involved in creating optimal matches between therapist and client. In some ways, similarity between client and therapist has been found to be advantageous, particularly for dimensions such as age, gender, and ethnicity (Beutler & Clarkin, 1990). Similarity is also likely to enhance the value placed on interpersonal treatment goals, friendship, and social recognition (Arizmendi, Beutler, Shanfield, Crago, & Hagaman, 1985; P. Talley, Strupp, & Morey, 1990). In contrast, dissimilarity between patient and client predicted better outcomes when therapists who valued a high level of autonomy worked with clients who had a high need for attachment and dependence. Conversely, therapists who were highly oriented toward attachment and dependency did better with clients who were highly self-sufficient and autonomous (N. Jacobson, Follette, & Pagel, 1986).

Beutler and his colleagues (Beutler, 1979; Beutler & Clarkin, 1990; Beutler et al., 2000; Harwood & Williams, 2003) have developed a model of treatment selection based primarily on the identification of relevant client characteristics. This approach relies on systematically identifying these characteristics and making recommendations based on empirically and clinically established relationships with treatment outcomes. These characteristics include degree of functional impairment, social support, level of problem complexity/chronicity, coping style, resistance, and subjective distress. This model, along with stages of change, is emphasized in the remainder of this chapter. The rationale for using this model is that it closely adheres to empirically validated research, uses many of the assessment techniques discussed in previous chapters, follows a clear sequence of decision making, and is comprehensive, while detailing a manageable number of variables.

The relevance and urgency of working with empirically validated methods of treatment planning are likely to significantly increase in the future. A powerful factor fueling this urgency is the current managed care movement, which will increasingly demand that both assessment procedures and interventions demonstrate their cost-effectiveness (Groth-Marnat, 1999; Groth-Marnat & Edkins, 1996; Groth-Marnat et al., 1995; Maruish, 2002). As a result, there is increasing pressure to demonstrate that assessment can quickly identify client problems, facilitate optimal treatment recommendations, and demonstrate the effectiveness of actual interventions. These "tools of the trade" must be able to provide these services in a way that has been demonstrated to be cost-effective. At the present time, the cost-effectiveness of assessment is not yet available but will most likely be forthcoming in the near future (Groth-Marnat, 1999, 2000b). Future research should clarify when assessment is and is not cost-effective and, in particular, demonstrate that assessment results can be used to save money by quickly and effectively developing a treatment plan, thereby avoiding misplaced and, possibly, ineffective or unnecessarily long treatment.

A SYSTEMATIC APPROACH TO TREATMENT SELECTION

When a practitioner is confronted with a client, relevant information needs to be acquired; and based on this information, a series of decisions and recommendations

should be developed. Beutler and his colleagues (Beutler & Clarkin, 1990; Beutler et al., 2000b; Beutler & Harwood, 2000; Harwood & Williams, 2003) have identified six patient dimensions and related these to different types of decisions (see Table 14.1). The first of these relates to *functional impairment* and has clear implications for general case management. Issues include the relative restrictiveness of therapy (inpatient/outpatient), whether medication should be considered, the intensity of treatment (duration and frequency), and what should be the immediate goals. The other five dimensions relate more to specific techniques of intervention than to general case management. Level of *social support* can be used to determine whether a client's social network can be relied on or whether it needs to be increased. The relative *complexity (and chronicity)* of a client's problem is important in considering whether the focus of treatment should be on specific, discrete, environmentally related symptoms, or more internal, chronic areas of conflict. In addition, *coping style* can help guide whether interventions should be on changing external behavior or directed at more internal insight-oriented levels of change, and level of *resistance* (reactance) has implications for how directive interventions should be. *Subjective distress* can be used to guide clinicians in determining whether the client's level of arousal should be increased or decreased. A final, seventh domain developed by Prochaska and DiClemente (1984, 1992) relates to tailoring interventions based on the *problem-solving phase* (stage of change) the client is in.

Each of these dimensions can be potentially assessed with a combination of formal tests, interview data, behavioral observations, and relevant history. This can range from a relatively short interview that focuses on each of the relevant domains, to an extensive battery consisting of a number of formal psychological tests. A rating scale (the *STS*

Table 14.1 Systematic steps in treatment planning

Variable	Treatment Considerations
1. Functional impairment	Restrictiveness (inpatient/outpatient) Intensity (duration and frequency) Medical vs. psychosocial interventions Prognosis Urgency of achieving goals
2. Social support	Cognitive behavioral vs. relationship enhancement Duration of treatment Psychosocial interventions vs. medication Possible group interventions
3. Problem complexity/chronicity	Narrow symptom focus vs. resolution of thematic unresolved conflicts
4. Coping style	Behavioral symptom oriented vs. internal insight oriented interventions
5. Resistance	Supportive, nondirective, or paradoxical vs. structured, directive interventions
6. Subjective distress	Increase/decrease arousal
7. Problem-solving phase	Understanding, exploration, and awareness vs. overt behavioral or interpersonal change

Clinician's Rating Form; D. Fisher, Beutler, & Williams, 1999) has also been developed to assist in summarizing the various ratings. In addition, a software program is available to provide narrative reports, project the course of treatment, graph a patient's relative standing on each of the assessed variables, assess various risks for the client, list most pressing problems, provide a series of brief minimanuals for each problem area, and evaluate outcome (*Systematic Treatment Selection: A Software Package for Treatment Planning;* Beutler & Williams, 1999). Completing the program takes between 20 and 40 minutes. Finally, clients (and clinicians) can enter data through a phone-in service so that the STS dimensions can be rated based on client responses and a treatment plan developed (see additional information at www.systematictreatmentselection.com or info@cbhti.com; Harwood & Williams, 2003). The following descriptions of these dimensions include a section on describing the construct followed by methods of assessment and different treatment implications based on the information derived from assessment. Relevant research to support important themes is cited but, given the often-immense volume of possible literature, it is not possible to provide an exhaustive listing of citations. Practitioners can use the following dimensions to organize their assessment procedures as well as to guide treatment interventions.

FUNCTIONAL IMPAIRMENT

A pressing problem related to any assessment is an evaluation of the severity of the problem. The core issue is to assess the extent to which the patient's problem interferes with his or her ability to effectively deal with everyday social, occupational, and intrapersonal requirements. This might have a direct relationship to the client's ability to cope, ego strength, level of insight, and chronicity of symptoms. In many cases, functional impairment relates to the extent to which the client is subjectively distressed. In many instances, however, subjective distress does not relate to the presence of severe problems. Examples include antisocial personalities who create suffering for others but do not feel particularly distressed themselves and schizoid personalities who are functioning on the fringes of society but do not feel particularly worried about their marginal status and level of dysfunction. The major distinction is that functional impairment is reflected in objective indicators of impairment. In contrast, subjective distress does not necessarily mean that the person is also impaired based on objective indicators.

There are numerous formal and informal assessment procedures for assessing functional impairment. Beutler and his colleagues (Beutler & Harwood, 2000; Gaw & Beutler, 1995) have summarized the relevant assessment dimensions to include the following:

- A problem that interferes with the client's ability to function during the interview.
- Poor concentration during assessment tasks.
- Distraction by minor events.
- General incapacity to function.
- Difficulty interacting with the clinician.
- Multiple impaired areas of performance in the client's daily life.

A Mental Status Examination is a structured means of obtaining useful information related to functional impairment.

One of the more useful psychometric indications of functional impairment is the presence of generally elevated scales on the MMPI-2/MMPI-A. Functional impairment is especially likely if elevations are found on scales on the right side of the profile (Paranoia, Schizophrenia, Hypomania). High Beck Depression Inventory-II (BDI-II) scores (30 or above) also suggest a high level of incapacity. Suicide level should always be assessed if the patient is depressed. Specific signs to alert the clinician to suicide risk are relevant critical items on the MMPI-2/MMPI-A (check critical items listed under Depressed Suicidal Ideation in Appendix H) or items 2 and 9 on the BDI-II. General elevations on the MCMI-III scales also suggest a high level of functional impairment, particularly if elevations occur on the Severe Personality Pathology or Severe Syndrome scales. The multiaxial *DSM-IV* (1994) system also provides methods for summarizing information relevant to estimating functional impairment. Impairment can be generally assessed by the specific type of diagnoses and is likely to be more severe if there are diagnoses on both Axis I and II and if there is the presence of severe disorders in the psychotic domain (schizophrenia, bipolar). In addition, the *DSM-IV* Global Assessment of Functioning rating specifically requests clinicians to provide an assessment of the level of functioning over the past year on a scale between 1 and 100.

Several noteworthy instruments covered in previous chapters (see Chapter 1) can also provide useful indicators of functional impairment. A high number of reported problems (*T* above 63) on the Brief Symptom Inventory (BSI; Derogatis, 1992) suggest high functional impairment as do high scores (*T* above 55) on the Trait Anxiety scale of the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983).

High Level of Functional Impairment

High levels of functional impairment have implications for the following five areas: restrictiveness of treatment, intensity of interventions (duration and frequency), use of medical/somatic versus psychosocial interventions, prognosis, and the urgency of achieving initial goals (see Beutler, Harwood, et al., 2003). Severe problems, particularly if the client is suicidal or cannot function in daily activities may require immediate inpatient care. Examples of diagnoses that may require inpatient care include bipolar mood disorders, psychotic conditions, major depression with suicidal intentions, acute substance abuse requiring detoxification, and some organic conditions that have resulted in significant decompensation. Initial treatment on an inpatient basis might later be reduced to partial hospitalization when the condition has become stabilized. Initial treatment for inpatients might need to be intensive. Outpatient interventions would be appropriate for the vast majority of clients whose problems are of mild to moderate severity (e.g., adjustment reactions, mild to moderate depression) and have greater resources.

The intensity of treatment (duration and frequency) varies from client to client based primarily on functional impairment. Greater duration of treatment is generally suggested for the following types of patients:

- Those with more serious diagnoses (e.g., borderline personality).
- Poor premorbid functioning.

- External stress seemingly of minor importance in the development and maintenance of the disorder.
- Age between 25 and 50 years.
- Client expectation that change takes time, and the technique used will be exploratory and insight oriented.
- Low level of social support.

In contrast, the following indicators suggest short duration of interventions:

- An acute disorder (e.g., adjustment disorder, acute reactive psychosis).
- External stress that seems to be of primary causal significance.
- Good premorbid level of functioning.
- Clients who expect change to occur quickly.
- Symptom-oriented focus of treatment, or crisis intervention.
- Structured, directive, and active interventions.
- Person who is either child/adolescent or elderly.
- High level of social support.

For some conditions, intermittent brief therapy throughout the life span at critical junctures might be an appropriate recommendation. At times, it might be appropriate to recommend no treatment, particularly if the person might have a negative response (e.g., some borderlines), no response (e.g., some antisocial personalities), spontaneous improvement (e.g., normal grief), or strongly respond to suggestions that he or she will improve rapidly with no treatment. Additional characteristics contraindicating psychotherapy might be a client's associating emotional pain with the change process, suspiciousness toward the therapist, and the client's need for control (Mohr, 1995).

Conditions such as schizophrenia, bipolar disorder, or severe anxiety states might require medical intervention (pharmacotherapy, electroconvulsive therapy) to enable clients to function well enough to become engaged in psychosocial or environmental interventions (see Sammons & Schmidt, 2001). Markers for such interventions might include poor orientation to time and place, poor short-term memory, marked confusion, clearly inappropriate mood, or low level of intelligence. Past clinical and research evidence has suggested severe and/or endogenous depression responds better to pharmacotherapy, whereas situationally caused mild and moderate depression responds better to psychosocial interventions. In contrast, the preponderance of current evidence indicates that both severe and endogenous depressions, as well as mild to moderate depression, can be treated at least as effectively with psychotherapy but without the potential for problematic side effects (Antonuccio et al., 1995; Free & Oei, 1989; Garvey, Hollon, & DeRubeis, 1994; McLean & Taylor, 1992; Simons & Thase, 1992). A clearer indication for antidepressant medication is a high number of vegetative symptoms (e.g., fatigue, insomnia, loss of appetite; Preston, O'Neal, & Talaga, 2002). Similar decision processes can be made for anxiety, psychotic, and bipolar disorders (see Preston et al., 2002).

To make prognostic judgments requires considering and integrating a diverse amount of information with particular reference to diagnosis, chronicity, subjective distress, and client resources (employment, abilities, social support). Research on prognosis is somewhat contradictory. On one hand, it might be argued that a person with a severe problem

will have difficulty overcoming it because it has progressed to such an extensive level. On the other hand, functional impairment may represent an extreme level in a fluctuating condition so that the person is likely to spontaneously return to an improved level of functioning. In addition, the potential magnitude of change is likely to be greater because the person has so much room for potential improvement. One guideline is that a high degree of psychiatric symptoms associated with the presence of somatic complaints (headaches, irritable bowel syndrome) is likely to suggest a poor prognosis (Blanchard, Schwarz, Neff, & Gerardi, 1988; Jacob, Turner, Szekely, & Eidelman, 1983). In contrast, patients presenting with severe levels of general anxiety and ambulatory depression typically do quite well with either psychosocial or pharmacological interventions (Elkin et al., 1989). Specific diagnosis can also be an important consideration because some diagnoses are likely to have poorer prognoses than others. For example, schizoid and antisocial personalities have difficulty engaging in productive therapy although certain Axis I conditions related to these personality types can often be targeted and effectively treated. It is a rule of thumb that the greater the chronicity of the disorder, the more difficult it is to treat. A final principle in prognosis is that clients with low levels of social support are not as likely to improve as those with high support (Billings & Moos, 1986; Moos, 1990).

Finally, severe problems suggest that the urgency of treatment is greater and should be focused around working with the symptomatic areas causing the client the greatest distress. Less severe problems mean that the urgency of change is less and the goals can change and be negotiated over time.

Low Level of Functional Impairment

In contrast to the previously described treatment considerations, low functional impairment suggests that treatment can be in an unrestricted setting (outpatient) and of relatively low frequency and duration. Psychosocial interventions will more likely be the predominant form of intervention, and there will be less urgency to rapidly define and achieve specific, symptom-oriented goals.

SOCIAL SUPPORT

Level of environmental support refers to the presence of a strong cohesive family and a secure form of employment. These external means of support can often modify the impact of other forms of stressors. High social support has also been associated with a favorable response to treatment (Mallinckrodt, 1996), as well as the ability to maintain the gains made through treatment (Zlotnick, Shea, Pilkonis, Elkin, & Ryan, 1996). Not only are the treatment gains higher for persons with high social support, but also they achieved these gains in a shorter period (Moos, 1990). In contrast, clients with low social support required more time to benefit from therapy.

Informal assessment of social support can be achieved by noting the following characteristics:

- The extent to which the client feels trusted and respected by the people in his or her life.

- The extent and quality of people he or she can confide in.
- Level of experienced loneliness.
- The extent he or she feels abandoned by family or friends.
- The extent to which the client feels a part of his or her family network.
- The number of friends the client has common interests with.

It should be stressed that assessing social support should not consider merely the number of people available for the person, but the quality the client feels regarding this support. It is one thing to be living with a large number of people and quite another to actually feel that it is possible to confide in those people.

There are also a number of more formal strategies for assessing social support. Probably the most frequently used scale is the Social Support Questionnaire (Sarason, Levine, Basham, & Sarason, 1983). Additional information related to social support might be the relative elevation of MMPI-2 scale 0 (Social Introversion). High scores suggest an inhibited, shy person who may find it difficult to have a large network of friends. In addition, elevations on 6 (Paranoia) and 8 (Schizophrenia) suggest that both the number, as well as the quality, of social support may be low. High scores on 1 (Hypochondriasis) and 3 (Hysteria) may indicate that, although the number of supports may be high, the quality of these supports may be poor. MCMI-III elevations may also provide useful information related to social support. High scores on Schizoid, Avoidant, Schizotypal, Paranoid, and Thought Disorder each might indicate both a low number, as well as low quality, of social support. Other scale elevations including Dependent, Histrionic, Narcissistic, Passive-Aggressive (Negativistic), Self-Defeating, and Borderline may have moderate to high social supports but these supports are also likely to be quite conflicted. For example, Dependents may have social supports but have achieved these supports through sacrificing their autonomy and sense of personal competence. They are also likely to experience some anxiety related to fears that this social support may not be permanent. The Narcissistic may similarly need extensive social supports but needs to manipulate others to maintain these supports, and there may be extensive hostility if there are any threats to the admiration he or she expects from others.

High Social Support

High social support suggests a shorter duration of therapy. Long-term intervention may even be contraindicated. Therapeutic gains are likely to occur relatively rapidly and be maintained. Therapies that enhance the quality of relationships are likely to be particularly effective, presumably because they are enhancing skills the person already has. In contrast, cognitive and behavioral therapies are likely to be less effective (Beutler et al., 2000).

Low Social Support

Low support suggests cognitive behavioral therapy is more effective than therapies designed to enhance relationships (Beutler et al., 2000). Both longer duration of therapy and the possibility of medication are indicated. It may be possible that a supportive group intervention would be useful in providing sufficient support to activate additional more relationship/interpersonal types of therapies.

PROBLEM COMPLEXITY/CHRONICITY

Some clients present with problems, such as simple phobias, that are narrow, focused, and either reinforced by or elicited by the environment. In contrast, other clients present problems of a diverse, complex nature. These problems are likely to be pervasive, enduring, and occur in many contexts. Instead of focusing on one or two specific behaviors, they involve diverse themes. A review of past relationships typically reveals that these themes have been enacted with persons in intimate relationships or who were in positions of authority. Examples might include passive-aggressive interactions with authority figures, conflicts between dependency and independence in intimate relationships, or consistently creating problematic relationships by choosing incompatible partners (e.g., alcoholics) despite the availability of more appropriate persons. These themes can be considered reenactments of internal, unresolved conflicts. While the overt goal of becoming involved with such relationships is to somehow resolve the conflicts and achieve a certain level of gratification, the result is usually further suffering. For these sorts of problems, the level of intervention needs to be quite different from problems that are narrow and symptomatic.

Problem complexity can be differentiated from functional impairment in several ways. Whereas *functional impairment* refers to level of dysfunction, *problem complexity* refers to underlying thematic patterns in the person's life that may or may not result in a high level of impairment. For example, a client may be functioning at a rather high level (low functional impairment) but still be quite troubled by chronic dissatisfactions in his or her relationships. These dissatisfactions may be the result of complex themes related to difficulties dealing with anger or issues related to dependency. Such themes may pervade not only one or two primary relationships, but most of the people the person comes into contact with. Whereas severe problems might be quite directly caused and reinforced by the environment (e.g., habits, reactions to stress), a complex problem is likely to be strongly related to internal unseen events. Furthermore, complex problems are likely to involve personality patterns that are spread across a wide variety of domains.

Problem complexity is more difficult to measure than most of the other factors relevant for treatment planning, in part because it is more theoretically bound. Clinicians from psychodynamic perspectives are far more likely to frame client difficulties as centering around symbolic, underlying, complex themes, whereas behaviorally oriented practitioners describe problems in narrower, concrete, environmentally-oriented language (Witteman & Koele, 1999). Although there is no clear resolution to this dilemma, three main features can be used to indicate problem complexity. The first is the presence of several problem domains or diagnoses (comorbidity), and the second is the presence of pervasive or recurrent patterns and themes of problem behaviors. A third feature suggesting a complex problem is the presence of a personality disorder or, at least, a personality style suggestive of a personality disorder. Beutler and his colleagues (Beutler & Harwood, 2000; Gaw & Beutler, 1995) have summarized indicators of problem complexity based on the following background information and behavioral observations:

- Behaviors are repeated as themes across unrelated situations.
- Behaviors are ritualized efforts to resolve underlying interpersonal or dynamic conflicts.

- Interactions seem primarily related to past rather than present relationships.
- Suffering rather than gratification is the result of the repetitive behavior.
- Problems are symbolic expressions of underlying unresolved conflicts.

In contrast, noncomplex problems are more often characterized as being:

- Situation-specific.
- Transient.
- Based on inadequate knowledge or skills.
- Having a direct relationship to initiating events.
- Stemming from chronic habits.

Another reason problem complexity is more difficult to assess is that there are no clear, well-defined instruments. However, some inferences can be made from existing tests. In particular, elevations on the MCMI personality scales are likely not only to suggest the presence of a complex problem, but also to provide information related to personality themes (Retzlaff & Dunn, 2003). The presence of a personality disorder as defined by *DSM-IV* (1994) criteria further suggests a complex problem. Additional information can be derived from themes noted in TAT story content or from the client's organization of his or her responses to the Rorschach. Both of these instruments can be quite useful in articulating how a client copes with his or her emotions, responds to stress, resolves conflicts, relates interpersonally, and defends against anxiety. Finally, the MMPI-2/MMPI-A can help clarify not only a client's symptom pattern, but also the dynamic interplay between the symptoms, coping strategies, likely patterns in interpersonal relationships, and overall personality structure. A chronic problem is indicated if Scales 1 (Hypochondriasis) and 2 (Depression) are both above 65 but Scale 1 is clearly higher (5 to 10 points or more) than 2. Problem chronicity is also suggested if both Scales 7 (Psychasthenia) and 8 (Schizophrenia) are above 65 but Scale 8 is clearly higher (5 to 10 points or more) than 7 (see Chapter 7).

High Problem Complexity

Complex problems are likely to respond best to broad treatments that are directed toward resolving long-standing underlying conflicts and changing patterns of interpersonal relationships. Depending on the problem, specific techniques might include:

- Two-chair work.
- Group or family therapy exploring patterns of responses.
- Dream work.
- Cathartic discharge.
- Enacting opposite patterns of how the client typically behaves.
- Exploring thematic patterns in behavior and relationships.
- Interpreting the transference.
- Interpreting resistance.
- Free association.

Low Problem Complexity

Noncomplex problems can be effectively treated by targeting specific symptoms, antecedents that elicit these symptoms, and consequences that maintain them. Depending on the problem, specific techniques might include:

- Behavioral contracting.
- Social skills training.
- Graded exposure.
- Reinforcement of target behaviors.
- Contingency management.
- Challenging dysfunctional cognitions.
- Practicing alternative cognitions.
- Practicing new self-statements.
- Self-monitoring.
- Paradoxical strategies.
- Counterconditioning.
- Relaxation.
- Deep muscle relaxation.
- Biofeedback.

COPING STYLE

Theory, research, and clinical observations indicate that client coping style varies on a continuum between externalization to internalization. Externalizers cope with their problems by impulsively acting out, externalizing blame, attributing the cause of their difficulties to bad luck or fate, and actively attempting to avoid their problems. They are not psychologically minded and, as a result, do not respond well to insight. In contrast, internalizers are more prone to blame themselves based in part on the perception that they do not have the sufficient skills or abilities to overcome their difficulties. Accordingly, they tend to experience more subjective distress than externalizers. To cope with this distress, they are likely to attempt to understand their difficulties in more depth.

Clinical indicators for externalization based on history and behavioral observations include the following (Gaw & Beutler, 1995):

- Projection.
- Blaming others for their problems.
- Paranoia.
- Low frustration tolerance.
- Extroversion.
- Unsocialized aggression.
- Manipulation of others.
- Distraction through seeking stimulation.
- Somatization with a focus on seeking secondary gains.

In contrast, internalizers are more likely to have the following characteristics:

- Introversion.
- Intellectualization.
- Constricted or overcontrolled emotions.
- Denial.
- Repression.
- Reaction formation.
- Minimizing difficulties.
- Social withdrawal.
- Somatization with symptoms related to the autonomic nervous system.

MMPI-2/MMPI-A assessment of externalization for clinical populations can be made by finding the sum of *T* scores on 4 (Psychopathic Deviance), 6 (Paranoia), and 9 (Mania) and then comparing this with the sum of *T* scores on the internalization measures of 2 (Depression), 7 (Psychasthenia), and 0 (Social Introversion). If the sum of externalization (4 + 6 + 9) is greater than internalization (2 + 7 + 0), the client can be considered an externalizer. Conversely, if the internalizing sum (2 + 7 + 0) is greater than the sum for externalization (4 + 6 + 9), the client is likely to internalize conflicts and stress (Beutler et al., 1991). Note that the preceding ratio has been designed for use with clinical populations who have at least some elevations on the MMPI-2/MMPI-A scales. For depressed patients, greater sensitivity can be achieved by calculating the sum of *T* scores for Scales 4 (Psychopathic Deviance) and 6 (Paranoia), which should be above 125 to fulfill the criteria for having an externalizing coping style.

Several additional measures might also provide useful information related to coping style. Low scores on the CPI socialization scale suggest an externalizing coping style, whereas high scores suggest a person who is more responsive and compliant (internal). The MCMI scales of Histrionic, Antisocial, Aggressive/Sadistic, and Paranoid conceptually suggest externalizing styles. In contrast, Avoidant, Depressive, Dependent, and Compulsive seem consistent with more internalizing styles of coping.

High Externalizers

Clients using externalizing coping strategies have better treatment outcomes when behavioral, symptom-oriented interventions, or specific techniques for building skills are used. In contrast, they do relatively poorly with techniques that attempt to enhance awareness and create insight (Beutler et al., 1991; Beutler & Clarkin, 1990; Beutler, Harwood, et al., 2003; Kadden, Cooney, Getter, & Litt, 1990). Techniques that are likely to be effective with externalizers include:

- Social skills enhancement.
- Assertiveness training.
- Group interventions.
- Anger management.
- Graded exposure.

- Reinforcement.
- Contingency contracting.
- Behavioral contracting.
- Questioning dysfunctional beliefs.
- Practicing alternate thinking.
- Stimulus control.
- Thought stopping.
- Counterconditioning.
- Relaxation.

High Internalizers

High internalizers benefit the most from techniques that emphasize the development of insight and the development of emotional awareness (Beutler et al., 1991; Beutler & Clarkin, 1990; Beutler, Harwood, et al., 2003; Kadden et al., 1990). Specific techniques might include:

- Cathartic discharge.
- Therapist-directed imagery.
- Dream interpretation.
- Direct instruction.
- Outside reading (bibliotherapy).
- Interpreting transference reactions.
- Interpreting resistance.
- Two-chair work.

RESISTANCE

Clients vary on the extent to which they are accepting and responsive to treatment versus being resistant and oppositional. This resistance is frequently a defense against what they perceive as others attempting to exert or intrude on their sense of control. Those who are most resistant are likely to have a constellation of traits including need for control, hostility, impulsivity, and direct avoidance (Dowd & Wallbrown, 1993). They may also have difficulty taking feedback and lack empathy. In addition to the preceding trait perspective, resistance can also be a state. The defensive or reactant state usually occurs when the client feels as if his or her freedom is somehow being threatened. Persons who are prone to be resistant are more likely to feel that they have a continual lack of personal control. As a result, they may compensate for this and establish a sense of control by acting in ways that oppose what is being requested or demanded of them. This is most likely to occur when the threatened area of freedom is important to the person and the individual making the request is doing so in an authoritative fashion such as through instruction, confrontation, directives, or structured techniques. Such a

structured, directive approach can potentially result in actual increases in client dysfunction. Understandably, highly reactant clients are likely to have a poorer prognosis than those who are more responsive and receptive.

Clinical indicators that may suggest high resistance include the following (Gaw & Beutler, 1995):

- Extreme need to maintain autonomy.
- Opposition to external influences.
- Dominance.
- Anxious oppositional style.
- History of interpersonal conflict.
- Poor response to previous treatment.
- Refusal to accept therapist interpretations.
- Incompletion of homework assignments.

In contrast, a low level of resistance is suggested by the following:

- Seeks direction.
- Submissive to authority.
- Open to experience.
- Accepts therapist interpretations.
- Agrees to and follows through with homework assignments.
- Indicates a tolerance to events beyond his or her control.

Although the MMPI-2/MMPI-A and MCMI-III do not have pure measures of resistance, elevations on some of the scales might be consistent with high resistance. Specifically, high scores on *L* and *K* are likely to have oppositional styles as would elevations on 6 (Paranoia) and possibly 1 (Hypochondriasis). Beutler et al. (1991) have used a combination of the MMPI research scales for anxiety (Taylor Manifest Anxiety Scale) and social desirability (Edwards Social Desirability Scale) as a measure of resistance. MCMI-III elevations on scales for Narcissistic, Negativistic (Passive/Aggressive), Paranoid, Aggressive/Sadistic, and Compulsive also suggest a defensive, oppositional person. In contrast, elevations on Dependent and Histrionic suggest a more responsive, compliant style. The most frequently used pure measure of resistance (reactance) is Dowd, Milne, and Wise's (1991) Therapeutic Reactance Scale with scores above 68 indicating sufficient resistance/reactance to have implications for treatment planning.

High Resistance

Strong empirical relationships have been found between positive treatment outcome and the use of nondirective, supportive, self-directed interventions for resistant clients (Beutler & Clarkin, 1990; Beutler, Moleiro, & Talebi, 2002; Beutler et al., 1991, 1996). Specific techniques might include:

- Self-monitoring.
- Therapist reflection.
- Support and reassurance.
- Supportive interpretation of transference.

In addition, paradoxical techniques have been found particularly effective with reactant clients and might include:

- Encouraging relapse.
- Prescribing that no change occur.
- Exaggeration of the symptom.

This is most likely to be true if resistance levels are quite high as might be reflected in scores above 84 (top 25%) on the Therapeutic Reactance Scale (Beutler et al., 1996; Debord, 1989; Dowd & Wallbrown, 1993; Horvath & Goheen, 1990).

Low Resistance

Clients who are responsive and compliant are likely to achieve the most gains when therapists use a more directive, structured approach (Beutler, Moleiro, & Talebi, 2002; Beutler et al., 1991, 1996; Gaw & Beutler, 1995; Horvath & Goheen, 1990). Specific techniques might include:

- Behavior contracting.
- Contingency management.
- Graded exposure.
- Direct hypnotic suggestion.
- Stimulus control.
- Cognitive restructuring.
- Developing alternative client self-statements.
- Directed imagery.
- Advice.
- Thought stopping.
- Therapist interpretation.

SUBJECTIVE DISTRESS

Subjective distress relates to the degree to which the person subjectively experiences his or her problem and is manifested primarily in heightened anxiety, confusion, or depression. A moderate level of subjective distress is useful because it motivates a client to become involved with change. It can lead to cognitive improvements including enhanced memory, faster performance, and higher intellectual efficiency. If a client's

distress becomes too high, however, it will be disruptive and result in deteriorated ability to function. The person then has difficulty appropriately processing information and concentrating. This interferes with the problem solving and behavioral experimentation required in therapy. A client whose level of subjective distress is too low will have difficulty becoming engaged in actively working to change behavior. Thus, there is an optimum window of distress that clinicians should try to achieve (Beutler & Harwood, 2000; Gaw & Beutler, 1995).

While there is some overlap with functional impairment and subjective distress, there are also a number of differences. As discussed previously, degree of functional impairment relates to objective indicators of poor functioning, whereas subjective distress is more an internal, subjective phenomenon. In addition, subjective distress can be quite changeable and may be controlled by environmental events. A client's level of subjective distress needs to be monitored from session to session or even within each session. A further contrast exists in the range and types of decisions relevant to either functional impairment or subjective distress. Issues relevant to functional impairment require wide-ranging decisions related to treatment setting (inpatient/outpatient), prognosis, treatment intensity (duration and frequency), and the general goals of intervention. The treatment implications of subjective distress are much narrower in that they provide guidance on whether arousal should be increased or decreased.

Frequent review of interview data, including behavioral observations and relevant history, is one of the best methods of monitoring a client's distress levels. Specific indicators of high distress include the following (Beutler & Harwood, 2000; Gaw & Beutler, 1995):

- Motor agitation.
- High emotional arousal.
- Poor concentration.
- Unsteady voice.
- Autonomic symptoms.
- Hyperventilation.
- Hypervigilance.
- Excited affect.
- Intense feelings.

In contrast, low levels of distress are indicated by:

- Reduced motor activity.
- Poor emotional investment in treatment.
- Low energy level.
- Blunted or constricted affect.
- Slow speech.
- Unmodulated verbalizations.
- Absence of symptoms.

MMPI-2/MMPI-A scales that are especially sensitive to subjective distress are *F,* 2 (Depression), and 7 (Psychasthenia). Collectively, these are frequently referred to as the "distress scales" (see descriptions under *F* scale, Scales 2 and 7, and the 27/72 code type in Chapter 7). However, motivation to change might be undermined if scales related to denial, resistance, and defensiveness are elevated (*L* and *K* as well as 3/Hysteria). A poor prognostic sign is a low 7 (Psychasthenia) with elevations on other scales suggesting psychopathology. This suggests that the client might be unrealistically relaxed regarding his or her difficulties or has given in to the inevitability of the problems. Additional measures of subjective distress are the Symptom Checklist 90-R, Brief Symptom Inventory (BSI), and the State-Trait Anxiety Inventory (STAI; see Chapter 13). A high level of distress is suggested if the Global Severity Index on the BSI is above 63 or the State Anxiety Score is in the top quartile.

High Subjective Distress

If subjective distress is quite high, an immediate goal is to reduce the anxiety level. This would be particularly urgent if the distress is sufficiently high to result in a significant disruption in the ability to cope. A wide variety of psychosocial techniques are available but are characterized by being supportive, structured, and designed to enhance relaxation. If a client's arousal is primarily expressed through physiological signs, techniques targeted at this level are warranted and might include the following:

- Progressive muscle relaxation.
- Hypnotically assisted physiological relaxation.
- Guided imagery.
- Biofeedback.
- Aerobic exercise.
- Graded exposure.

Arousal that is more socially or cognitively related might be most effectively reduced through the following techniques:

- Meditation.
- Reassurance.
- Emotional support.
- Cathartic discharge.
- Supportive challenging of dysfunctional cognitions.
- Time management.
- Thought stopping.

Pharmacotherapy might be useful but should be accompanied by learning new coping skills so that medication can be discontinued as soon as possible. The newly acquired coping skills then decrease the likelihood of relapse after the medication has been discontinued.

Low Subjective Distress

Clients with low subjective distress are likely to be associated with involuntary referrals. Experiential strategies can confront clients with the impact and consequences of their difficulties and are likely to increase distress to a level that makes them more open to changing their behavior. Possible techniques are:

- Two-chair work.
- Symptom exaggeration.
- Experiential role plays.
- Confrontation.
- Family therapy initially focusing on the impact of client behavior on family members.
- Overt practice.
- Predicting the recurrence of symptoms.
- Discussing painful memories.
- Accessing affective responses.
- Directed imagery.
- Interpretation of the transference.
- Interpretation of resistance.

PROBLEM-SOLVING PHASE

Clients undergo a series of steps during the process of change. Accordingly, any client referred for evaluation may be at a different stage in the change process. Some individuals might be simply considering the possibility of change but have not yet struggled with the specifics of how to accomplish it. This might be particularly true for involuntary referrals who are resistant and experiencing a low level of subjective distress. On the other extreme might be a client who has already taken a number of clear steps for change but is seeking help to prevent relapse. According to the stage of change, a client might require somewhat different approaches. However, considering stage of change may not be relevant for disability, medical, or many court assessments (e.g., personal injury) because facilitating change may not be part of the referral question. In these cases, assessment of the current level of functioning or differential diagnosis becomes the focus of the report.

The stages of change are likely to be quite variable. One person might pass through the different stages quite rapidly and another who is perhaps more ambivalent or less directed might have been considering the possibility of change for years. During the process of successful therapy, it would be expected that the client would have undergone all the different stages at some point. As a result, practitioners need to be continually aware of possible changes in the stage of change and adapt their interventions accordingly. In addition, a client might have several problem areas, especially if the problem is complex, and each area might be at a different stage in the change process. This variability requires a flexible approach depending on which area is being addressed.

Prochaska and DiClemente (1984, 1992) have described the following five stages in the change process: precontemplation, contemplation, preparation, action, and maintenance. Each stage has a different set of tasks that must be accomplished before proceeding to the next stage. The first three stages are processes that occur before any actual change or actual attempts at concrete change. In the *precontemplation* stage, people have little intention of changing behavior or attitudes. They might be vaguely aware that change needs to occur but, for the most part, they are unaware of the possible importance of change. In contrast, other people they relate with can clearly see the need for change. As a result, these clients are likely to be referred or seek treatment when the legal-justice system threatens to punish them, a spouse threatens to leave them, parents threaten to disown them, or an employer threatens to dismiss them. Under these conditions, change is likely to proceed only if there is either continual outside pressure, or the actual client internalizes the need for change. When individuals begin to more seriously consider change, they can be considered to be in the *contemplation* stage. At this point, they are aware that they have a problem and are concerned with how coping with the problem might best be accomplished. However, they have not yet committed themselves to the process. In the *preparation* stage, they have become more committed to change, which is represented by their intent to take action in the near future. This intent may also be accompanied by the possible presence of minor experiments with new behaviors. Because they are not yet clear on how best to accomplish their intended change, they may need help considering all relevant options and choosing the optimal strategy for implementing the change.

The final two steps in the change process focus on actually implementing the change and ensuring that it is maintained. *Action* is the point at which clients actually change their environment, attitudes, or behavior. Often, this requires a considerable amount of time and energy and, as a result, individuals must be highly committed. Changes at this point are most clearly visible to others. The preceding preparatory and contemplative processes should not be underestimated, however, because they are crucial in determining the relative success of any change. During the *maintenance* stage, individuals work to consolidate change and prevent relapse.

The following interview questions can help determine the stage of change: Do you intend to change in the near future? Are there current changes you are going through? Have you made changes? Are you currently working to prevent relapse? These questions might also be incorporated into an intake form (Prochaska, Norcross, & DiClemente, 1994). It may be necessary to probe or otherwise obtain clarification to clearly determine the stages of change. Formal assessment of the stages of change can also be made on the 32-item Stages of Change Scale (McConnaughy, Prochaska, & Velicer, 1983).

Research has generally supported the clinical utility and predictive validity of tailoring interventions according to the different stages of change. This research has primarily focused on problems such as addictive behaviors, weight control, eating disorders, sunscreen use, and exercise acquisition (Geller, Cockell, & Drab, 2001; Prochaska, 2000; Prochaska, DiClemente, & Norcross, 1992; Prochaska, Rossi, & Wilcox, 1991). Further research needs to be conducted to determine its applicability for a wider range of problem areas (Whitelaw, Baldwin, Bunton, & Flynn, 2000). In the areas researched, there is generally evidence that tailoring interventions toward

the stage of change can optimize treatment outcome (Petrocelli, 2002; Prochaske, 2000; Prochaska et al., 1992).

- *Precontemplation Stage.* This is often, although not necessarily, consistent with involuntary referrals. As a result, resistance level may be high and subjective distress low, such that interventions would need to be made accordingly (e.g., increase arousal; use nondirective, supportive techniques; paradoxical interventions). Because these clients might feel ambivalent about treatment, it is crucial to spend time building rapport and discussing areas that work or don't work in their lives.
- *Contemplation and Preparation Stages.* As in the previous stage, enhancing the relationship is particularly important. Providing understanding and awareness is also crucial. This should include exploring the interpersonal or behavioral patterns of the client, reasons for and against changing, and the different strategies for creating change. An inventory of client strengths or resources and weaknesses might also be useful. The first three stages might be most consistent with humanistic or psychodynamic approaches that stress insight, exploration, value clarification, novel experiences, and clarification of personal goals.
- *Action Stage.* A wide variety of specific, concrete techniques might be used. The selection of these techniques depends in part on areas such as functional impairment, problem complexity, subjective distress, and resistance. Specific strategies can be implemented that might involve changes in concrete behavior, patterns of interpersonal relationships, self-statements, or ways of experiencing the world. Cognitive or behavioral techniques might be most effective at this point, particularly stimulus control, graded exposure, cognitive restructuring, role plays, social skills training, or counterconditioning.
- *Maintenance Stage.* At this point, the therapist can become like a coach or a consultant who advises and encourages the client. A crucial consideration is how relapse is most likely to occur and to develop countermeasures to prevent these situations from occurring or at least to minimize their impact over a longer period. Specific techniques might include stimulus control, social contracting, enhancing social support, anger management, or a behavioral contract requiring the person to take preventive measures if relapse seems likely.

The preceding seven dimensions are intended to be logically consistent as well as manageable. The use of the model might be particularly crucial during training for new clinicians or skill enhancement of more experienced ones. With practice, it is likely that many of the features will become progressively more internalized, perhaps requiring less formal assessment. A briefer, more clinical assessment of the dimensions may also be required when short-term interventions (e.g., crisis intervention) are the only options available.

As further research provides more precise definitions of empirical relationships, additional dimensions will likely be included. There may also be further integration with both differential therapeutics and therapist-client matching. Each of these developments will bring clinicians closer to Paul's (1967) previously stated ultimate goal of combining the best treatment with the optimal mix of therapist, client, problem, and context.

RECOMMENDED READING

Beutler, L. E., & Groth-Marnat, G. (2003). *Integrative assessment of adult personality* (2nd ed.). New York: Guilford Press.

Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive psychotherapy: A practical guide to systematic treatment selection.* New York: Oxford University Press.

Groth-Marnat, G., Roberts, R., & Beutler, L. E. (2001). Client characteristics and psychotherapy: Perspectives, support, interactions, and implications. *Australian Psychologist, 36,* 115–121.

Jongsma, A. E., & Peterson, L. M. (1995). *The complete psychotherapy treatment planner.* New York: Wiley.

Norcross, J. (Ed.). (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient's needs.* New York: Oxford University Press.

Chapter 15

PSYCHOLOGICAL REPORT

The psychological report is the end product of assessment. It represents the clinician's efforts to integrate the assessment data into a functional whole so that the information can help the client solve problems and make decisions. Even the best tests are useless unless the data from them is explained in a manner that is relevant and clear, and meets the needs of the client and referral source. This requires clinicians to give not merely test results, but also interact with their data in a way that makes their conclusions useful in answering the referral question, making decisions, and helping to solve problems.

An evaluation can be written in several possible ways. The manner of presentation used depends on the purpose for which the report is intended as well as on the individual style and orientation of the practitioner. The format provided in this chapter is merely a suggested outline that follows common and traditional guidelines. It includes methods for elaborating on essential areas such as the referral question, behavioral observations, relevant history, impressions (interpretations), and recommendations. This format is especially appropriate for evaluations that are problem oriented and that offer specific prescriptions for change. Additional alternatives for organizing the report are to use a letter format, give only the summary and recommendations, focus on a specific problem, summarize the results test by test, write directly to the client, or provide client descriptions around a particular theory of personality. The sample evaluations vary somewhat from the suggested format, although they usually still include the essential categories of information that are discussed in this chapter.

One general style to avoid is sometimes referred to as a "shotgun" report (Tallent, 1992, 1993). This provides a wide variety of often-fragmented descriptions in the hope that some useful information can be found within. The shotgun approach is usually vague, stereotyped, and overinclusive. The recommendations for treatment are often neither specific nor practical. The most frequent reason for a shotgun report is a referral question that is too general, vague, and, therefore, poorly understood. In contrast, the "case-focused" report centers on the specific problems outlined by the referring person. It reveals unique aspects of the client and provides specific accurate descriptions, rather than portraying stereotypes that may also be overly "theory linked" or "test linked." Furthermore, the recommendations for treatment are both specific and practical. The general approach of the case-focused report is not so much *what* is to be known, but rather *why* different types of information are important for the purposes of the report.

The creation of a case-focused report involves understanding and applying several basic principles. First, the report should use action-oriented language rather than

metapsychological abstractions. This means the client's ongoing behaviors and likely personality processes should be described in relation to different situations. The use of "action-language" links the person with specific behaviors, forces reports to address specific therapeutic issues, and conveys a better understanding of the client's active role in the testing situation. Second, the recommendations in a case-focused report need to directly relate to what specifically can be done for this client in his or her particular environment. They may apply to areas such as occupational choice, psychotherapy, institutional programs, or additional evaluation. In certain types of referrals, however, especially clients self-referred for psychotherapy, an important goal may be to help them increase their level of personal insight. In these cases, a wider description of the client that includes a number of different topics might be more appropriate than the narrower, problem-solving approach. In addition, there should be a focus on that which differentiates one person from another. This means avoiding discussions of what is average about the client, and emphasizing instead what stands out and is unique to this individual. Further, there is a current trend, consistent with the case-focused approach, toward de-emphasizing diagnosis and etiology. There is, rather, an emphasis on current descriptions of the person that are tied to specific behaviors. In certain cases, especially in a medical setting, the clinician may still need to provide diagnoses in addition to behaviorally oriented descriptions. Another consideration is that a case-focused report should be written with an awareness of the point of view of the intended readers. This includes taking into consideration their level of expertise, their theoretical or professional orientation, the decisions they are facing, and the possible interpretations they are likely to make of the information.

A final point is that the quality and usefulness of a report is typically enhanced if the practitioner is knowledgeable about the area or type of issue the client is experiencing. Such knowledge helps to increase the depth of the interpretations and provides relevant information or a general "map" of the problem area that can be used to help ensure that all relevant aspects have been covered. Importantly, background knowledge on the problem area provides relevant information on a range of interventions as well as the effectiveness of these interventions. For example, knowledge regarding depression means that the practitioner is aware of its causes, variety of ways in which it is expressed, options for interventions, and when further assessment is indicated (for suicide potential). Often consulting a well-written, up-to-date chapter will provide sufficient information. In the general clinical area, useful resources are Barlow's (2001) *Clinical Handbook of Psychological Disorders* (4th ed.), R. Meyer's (1996) *The Clinician's Handbook* (3rd ed.), or Kaplan and Sadock's (2001) *Kaplan & Sadock's Pocket Handbook to Clinical Psychiatry* (3rd ed.). Persons preparing neuropsychological reports might consult Groth-Marnat's (2000a) *Neuropsychology Assessment in Clinical Practice: A Guide to Test Interpretation and Integration,* Lezak's (1995) *Neuropsychological Assessment,* or Snyder and Nussbaum's (1998) *Clinical Neuropsychology: A Pocket Handbook for Assessment.* Educational report writers might benefit from reading relevant sections in Sattler's (2001, 2002) *Assessment of Children* or Walker and Roberts' (2001) *Handbook of Clinical Child Psychology.* A particularly useful resource when doing vocational assessments is Lowman's (1991) *The Clinical Practice of Career Assessment: Interests, Abilities, and Personality.*

GENERAL GUIDELINES

Length

The typical psychological report is between five and seven single-spaced pages (Finn, Moes, & Kaplan, 2001). However, the length can vary substantially based on the purpose of the report, context, and expectations of the referral source. In medical contexts, a two-page report is not uncommon. This parallels the format of many physician reports that are a similar length. In contrast, legal contexts often require reports that are from seven to ten pages because of the greater need for documentation combined with more extensive referral questions. It is not unusual for a psychologist serving as an expert witness to not only evaluate a client, but also anticipate and defend himself or herself against rebuttals as well as comment on reports made by other mental health professionals. The more moderate (and frequent) five- to seven-page report is particularly prevalent in psychological, educational, and vocational contexts.

Style

The style or "flavor" of a report is influenced primarily by the training and orientation of the examiner. The clinician can choose from four general report-writing approaches: literary, clinical, scientific, and professional (Ownby, 1997; Tallent, 1992, 1993). Each style has unique strengths, and all have a number of liabilities. The literary approach uses everyday language, is creative, and often dramatic. Although it can effectively capture a reader's attention and provide colorful descriptions, it is often imprecise and prone to exaggeration.

The clinical approach focuses on the pathological dimensions of a person. It describes the client's abnormal features, defenses, dynamics involved in maladjustment, and typical reactions to stress. The strength of the clinical approach is that it provides information about areas in need of change and alerts a potential practitioner to likely difficulties during the course of treatment. However, such a report tends to be one-sided in that it may omit important strengths of the person. The result is likely to be more a description of a "patient" than a person. Such a maladjustment bias is a frequent difficulty in clinical psychology and results in a distorted, unrealistic view of the client. Although most clinical reports should describe a person's problem areas, these problem areas should be given appropriate emphasis in the context of the client's relevant strengths and resources.

The scientific approach to report writing emphasizes normative comparisons, tends to be more academic, and, to a lesser extent, relates to the nature of a client's pathology. The scientific style differs from the other two approaches chiefly in its reference to concepts, theories, and data. It looks at and describes test findings in an objective, factual manner. Thus, there might be frequent references to test data, normative comparisons, probability statements, and cutoff scores to be used for decision making. A scientific approach is likely to discuss the person by addressing different, often isolated, segments of personality. Thus, areas such as a client's cognitive, perceptual, and motivational abilities may be described as discrete and often unrelated functions. Although the scientific

approach is objective and factual, it has been criticized for violating the unity of personality. Many readers, particularly those from other disciplines, do not respect or empathize with scientific evaluations and perceive them as cold, distant, and overly objective. Purely data-oriented evaluations can potentially do the profession a disservice by reinforcing the view that an assessment is like a laboratory test rather than a professional consultation with a clinician. Furthermore, a focus on factual data may not address the practical decisions the client and referral source are facing.

In actual practice, it is unusual to find a pure example of a literary, clinical, or scientific report. Clinicians generally draw from all three approaches but typically emphasize one. An important part of effective report writing is the ability to evaluate the assets and limitations of each style, and to maintain a flexible orientation toward appropriately combining them. In any one report, there may be a need to use creative literary descriptions, elaborate on different pathological dimensions, or provide necessary scientific information. Again, the key is to avoid the pitfalls associated with specializing in any one of these styles and to emphasize instead their relative strengths.

Ownby (1997) stresses that the most important style to use in report writing is what he refers to as a *professional style.* This is characterized by short words that are of common usage and that have precise meanings. Grammatically, writers should use a variety of sentence constructions and lengths to maintain the reader's interest. The paragraphs should be short and should focus on a single concept. Similar concepts should be located close to one another in the report. Whereas Hollis and Donna (1979) urge writers to use short words, short sentences, and short paragraphs, the *Publication Manual of the American Psychological Association* (5th ed., 2001) recommends varying the lengths of sentences and paragraphs. The result should be a report that combines accuracy, clarity, integration, and readability.

Presenting Test Interpretations

Clinicians generally prefer to orient their reports around specific hypotheses or different relevant domains, or adhere to interpreting the data test by test. The hypothesis-oriented model focuses heavily on answering specific questions asked by the referral source. The report tends to be highly focused, well integrated, and avoids any extraneous material. For example, if a referral source asks whether person X is brain-damaged, all the interpretations based on the test data are directed toward answering whether this hypothesis is supported.

A domain-oriented report discusses the client in relation to specific topics such as cognitive abilities, interpersonal relationships, vocational abilities, or sexuality. This approach is comprehensive, indicates the client's strengths and weaknesses, and typically gives the reader a good feel for the person as a whole. The referral question is still answered but is addressed by responding to specific domains relating to the referral question. Readers tend to prefer and better comprehend integrated reports written by addressing functional domains rather than test scores. The weakness of domain-oriented reports lies in the potential to provide too much information, thus overloading the reader.

Occasionally, a report is organized by presenting the results of each test, one at a time (WAIS-III, Bender, MMPI-2, etc.). This approach clarifies the source of the data and enables the reader to understand more clearly how the clinician made his or her inferences.

It is also relatively easy for the examiner to organize the results. These advantages are offset by significant disadvantages. The emphasis on tests can distract the reader and tends to reduce the client from a person to a series of test numbers. This is reflected in that readers of reports, regardless of their theoretical or disciplinary background, do not respond well to this style of report writing (Mendoza, 2001; Tallent 1992, 1993). A test-by-test presentation also reflects a failure to integrate the data. This may be a particular issue because it is not uncommon for inconsistencies to occur among different test scores. Often, only half of all possible interpretations listed in an interpretive manual or computer narrative are actually true for a particular client. It is up to the clinician to determine which of these do or do not apply for the person. Sometimes a report writer using a test-by-test approach hedges his or her "interpretations" by using a phrase such as "Other persons with similar test profiles have the following qualities . . ." The referral source, however, doesn't want to know about *other* people but is concerned with *this* client, at this time, living in a certain context. A test-by-test interpretation, then, suggests that the practitioner has neither adequately conceptualized relevant dynamics, nor fully understood the area under investigation (Mendoza, 2001; Wolber & Carne, 1993). It also encourages the belief that an examiner is a technician who merely administers tests rather than a clinician who uses multiple sources of information to answer referral questions and help people solve problems they are facing. The existing literature is unanimous in discouraging a test-by-test style and, instead, strongly recommends an integrated, case-focused, problem-solving style (Beutler & Groth-Marnat, 2003; Groth-Marnat, in press; Kvaal, Choca, & Groth-Marnat, 2003; Mendoza, 2001; Sattler, 2002; Tallent, 1992, 1993; Wolber & Carne, 1993; Zuckerman, 2000).

Topics

There is an extremely wide range of topics or domains that clinicians may decide to discuss in their reports. These topics serve as conceptual tools that enable report writers to give form and direction to the information they are trying to communicate. The three most common topics are likely to be related to cognitive functioning, emotional functioning (affect/mood), and interpersonal relations. Many reports can be adequately organized around these three domains. Additional topics include personal strengths, vocational aptitudes, suicidal potential, defenses, areas of conflict, behavior under stress, impulsiveness, or sexuality. Often, an adequate case-focused report can be developed by describing just a few of these topics. For example, a highly focused report may elaborate on one or two significant areas of functioning, whereas a more general evaluation may discuss seven or eight relevant topics. Table 15.1 is a representative list of topics that may be considered for inclusion in an evaluation. This list is by no means complete but can provide a general guide for the wide range of possible topics from which a report writer can choose.

Deciding What to Include

The general purpose of a psychological evaluation is to provide information that will be most helpful in responding to the referral question and meeting the needs of the client. In this context, the clinician must strike a balance between providing too much

Table 15.1 Examples of general topics around which a case presentation may be conceptualized

Achievement	Emotional functioning	Psychopathology
Affect	Fixations	Rehabilitation needs
Aggressiveness	Flexibility	Rehabilitation prospects
Antisocial tendencies	Frustrations	Resistance
Anxieties	Functional impairment	Sentiments
Aptitudes	Goals	Sex
Attitudes	Hostility	Sex identity
Aversions	Identity	Sex role
Awareness	Intellectual controls	Significant others
Behavioral problems	Intellectual levels	Situational factors
Biological risk factors	Interests	Social consequences of
Cognitive functioning	Interpersonal relations	behavior
Cognitive skills	Interpersonal skills	Social role
Cognitive style	Irrational cognitions	Social structure
Competency	Lifestyle	Social support
Conflicts	Mood	Special assets
Content of consciousness	Needs	Strengths
Coping style	Outlook	Subjective feeling states
Defenses	Perception of environment	Symptoms
Deficits	Perception of self	Treatment prospects
Developmental factors	Personal consequences of	Value system
Diagnostic considerations	behavior	Vocational topics
Drives, dynamics	Placement prospects	
Emotional controls	Problem complexity	

Adapted from N. Tallent (1988). *Psychological Report Writing* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall, p. 120.

information and providing too little, and between being too cold and being too dramatic. As a rule, information should be included only if it serves to increase the understanding of the client. For example, descriptions of a client's appearance should be oriented toward areas such as his or her level of anxiety or resistance. A client might be described as hesitant in his or her approach to tasks and may say something like, "Why do I have to take all these tests anyway?" If the person was dressed in bizarre clothes and his or her hair was unkempt or dyed purple, this information might also be quite important to include. Generally, however, information regarding the types of clothing the person is wearing or color of his or her eyes or hair is not relevant.

The basic guidelines for deciding what to include in a report relate to the needs of the referral setting, background of the readers, purpose of testing, relative usefulness of the information, and whether the information describes unique characteristics of the person. After these general guidelines have been considered, the next step is to focus on and organize the information derived from the tests. For example, if a general review of aspects of personality is the purpose of the report, a clinician can look at each test to determine what information it can provide.

A further general rule is that information should focus on the client's unique method of psychological functioning. A reader is concerned not so much with how the client is

similar to the average person as in what ways he or she is different. A common error in psychological reports is the inclusion of generalized statements that are so vague, they could apply to the majority of the population. These vague, generalized statements are likely to be unconditionally accepted as applying to a person even though they are randomly selected. For example, Sundberg (1955) administered a "personality" test to a group of students and gave them all identical "interpretations" based on universal or stereotyped personality descriptions composed of 13 statements, such as:

- You have a great need for other people to like and admire you.
- You have a tendency to be critical of yourself.
- You have a great deal of unused capacity you have not turned to your advantage.
- While you have some personality weaknesses, you are generally able to compensate for them.
- At times, you have serious doubts as to whether you have made the right decision or done the right thing.

Virtually all students used in the study reported that the evaluation statements were accurate descriptions of themselves. Other studies suggest that, not only were students unable to discriminate between fictitious and genuine feedback, but they may even have preferred generalized fictitious results, particularly if they were framed within a positive context (Dies, 1972). This uncritical acceptance of test interpretations might be even further encouraged when objective-appearing, computer-generated interpretations are used (Groth-Marnat & Schumaker, 1989). W. Klopfer (1960) has referred to this uncritical acceptance of universally valid statements as the "Barnum effect," in reference to Phineas Barnum's saying, "There is a fool born every minute." Although "universal statements" may add to the "subjective" validity of the report when read by the client, such statements should be avoided in favor of stressing the person's essential uniqueness.

After the data, conclusions, and recommendations have been outlined, the next step is to decide on the manner in which to present them. This involves clear communication about the relative degree of emphasis of the results, type of report, proper use of terminology, and the extent to which the raw data will be discussed.

Emphasis

Careful consideration should be given to the appropriate emphasis of conclusions, particularly when indicating the relative intensity of a client's behavior. General summaries may be given, such as "this client's level of depression is characteristic of inpatient populations," or the relative intensity of certain aspects of a client's disorder may be more specifically discussed. To continue with the example of depression, a clinician may discuss the client's cognitive self-criticisms, degree of slowed behavior, extent of social support, level of social skills, or suicidal potential. In addition to discussing and giving the appropriate degree of emphasis to a client's pathology, his or her psychological strengths need to be compared with his or her relative weaknesses. Furthermore, the report should not discuss areas of minor relevance unless they somehow relate to the purpose of the evaluation. To achieve proper emphasis, the examiner and the referral source

must clarify and agree on the purpose of the evaluation. Only after this has been accomplished can the examiner decide whether certain information should be elaborated in-depth, briefly mentioned, or deleted.

When clinicians present their conclusions, it is essential that they indicate their relative degree of certainty. Is a specific conclusion based on an objective fact, or is the clinician merely presenting a speculation? For example, the statement "John scored in the dull normal range of intelligence" is an objective fact. However, even in this case, examiners may want to give the standard error of measurement to provide an estimate of the probable range of scores. If only mild supporting data is available or if clinicians are presenting a speculation, phrases such as "it appears . . .", "tends to . . .", or "probably . . ." should be used. This is especially important when clinicians are attempting to predict a person's behavior, because the predicted behavior has not yet been observed. It may be useful for clinicians to indicate that their predictions cannot be found directly in the tests themselves, but rather represent inferences that have been made based on the test data. There should be a clear distinction between what the client did, and what he or she anticipates doing. If a statement made in a report is a speculation, it should be clearly indicated that the statement has only a moderate or small degree of certainty. Whenever a speculation is included, it should be relevant to the referral question.

Improper emphasis can reflect an incorrect interpretation by the examiner, and this misinterpretation is then passed down to the reader. Clinicians sometimes arrive at incorrect conclusions because their personal bias results in selective perception of the data. Thus, clinicians can develop an overly narrow focus with which they overlook potentially relevant data. Personal bias may result from factors such as a restrictive theoretical orientation, incorrect subjective feelings regarding the client, or an overemphasis on pathology. Inaccurate conclusions can also result from attempts to please the referral source or from interpretations based on insufficient data. The reader may also be likely to misinterpret the conclusions if the report is generally overspeculative or if speculations are not specified as such but, rather, are disguised as assertions. If speculations are overly assertive, this may not only lead the reader to develop incorrect conclusions, but also the report may become overly authoritative and dogmatic, perhaps leading readers to become irritated and skeptical.

Misinterpretations can also result from vague and ambiguously worded sentences that place incorrect or misleading emphasis on a client's behavior. A statement such as "the client lacks social skills" is technically incorrect because the client must have some social skills, although these skills may be inadequate. A more correct description would be to state that the client's social skills are "poorly developed" or "below average." Likewise, a statement such as "the client uses socially inappropriate behavior" is subject to myriad interpretations. This could be rephrased to include more behaviorally oriented descriptions, such as "frequently interrupts" or "would often pursue irrelevant tangents."

One technique of emphasizing results is to place the most relevant sections in boldface or italics. For example, the major identified symptoms, most important findings, and the major recommendations could all be placed in boldface. This enables persons reading the report to more easily absorb the most salient features. In addition, they can easily relocate major points that have been made. However, this technique should be used sparingly because readers can become easily saturated with too much boldface

print. This means that, instead of placing entire paragraphs or sentences in boldface, only key phrases should be given this method of emphasis.

The areas, extent, and method of emphasis significantly contribute to the conclusions of a report. However, responsibility for a report's conclusions rests on the clinician. This responsibility should not and cannot be transferred to the tests themselves. To take this a step further, decisions made about a person should never be in the hands of tests, which may even have questionable validity in certain contexts. Rather, conclusions and decisions regarding people should always be in the hands of responsible persons. Thus, the style of emphasizing results should reflect this. Phrases such as "test results indicate . . ." may give the impression that the examiner is trying to hide behind and transfer responsibility for his or her statements onto the tests. Not only is this not where the responsibility should be, but the reader may develop a lack of confidence in the clinician. If clinicians feel uncertain about a particular area, they should either be clear about this uncertainty or, if they cannot personally stand by the results, exclude the results from the report.

Use of Raw Data

When writing the impressions and interpretation section, a report writer should generally avoid adhering too closely to the raw data. For certain purposes, however, it may be useful to include raw data or even to describe the tests themselves. Test descriptions allow untrained persons to know specific behaviors the client engaged in rather than merely the final inferences. As a result, consumers of reports rate the inclusion of behavioral referents quite favorably (Finn et al., 2001). For example, a report may include a description such as "Mr. A had an average level of recall for short-term visual information, as indicated by his being able to accurately recall and reproduce five out of a possible nine geometric designs that he had previously worked with for five minutes." This sentence provides a more behaviorally referenced description than one like "Mr. A had an average level of recall as measured on the Bender memory." Thus, a test description is apt to give the reader a more in-depth, precise, and familiar reference regarding the subject's abilities. In addition to the test descriptions themselves, test responses can serve to make the description behavior specific and to balance high-level abstractions with concrete responses. For example, a clinician might discuss a client's impulsiveness and include illustrative items on the MMPI-2 (items adapted from actual MMPI-2 items), such as:

> During one period when I was a child, I used to shoplift. (True)
>
> At times, I have found it nearly impossible to keep from stealing. (True)

In discussing the same issue, a clinician could also include a portion of a TAT story that illustrates a similar point:

> . . . so he took the violin and, without even thinking about it, threw it into the fire and ran.

It is crucial to stress that the purpose of providing raw data and behavioral descriptions is to enrich and illustrate the topic and not to enable the reader to follow

the clinician's line of reasoning or document the inferences that have been made. In developing inferences, clinicians must draw on a wide variety of data. They cannot possibly discuss all the patterns, configurations, and relationships they used to come to their conclusions. Any attempt to do so would necessarily be overly detailed, cumbersome, and incomplete. Statements such as, "In considering the pattern of elevated Scales 4 and 9 on the MMPI, it is safe to conclude . . ." are unnecessary and rarely contribute to a report's overall usefulness. In certain types of reports, such as those for legal purposes, it might be helpful to include some raw data, not so much to repeat the thinking process of the clinician, but more to substantiate that the inferences are data based, to provide a point of reference for discussing the results, and to indicate what assessment procedures were used.

Terminology

Several arguments have been made in determining whether to use technical or nontechnical language in psychological reports. It might be argued that technical terminology is precise and economical, increases the credibility of the writer, and can communicate concepts that are impossible to convey through nontechnical language. However, a number of potential difficulties are often encountered with the use of technical language. One of the more frequent problems involves the varying backgrounds and levels of sophistication of the persons reading the report. The most frequent readers of reports include teachers, administrators, judges, attorneys, psychiatrists, nonpsychiatric physicians, and social workers. Increasingly, the clients themselves have access to and read the reports (Harvey, 1997). Thus, many, if not most, consumers of reports do not have the necessary background to interpret technical terminology accurately. Even psychologists with different theoretical persuasions may be apt to misinterpret some of the terms. Take, for example, the differing uses of *ego* by Freud, Jung, and Erikson. Also, the term *anxiety* might have several different categories of use. Although technical words can undoubtedly be precise, their precision is helpful only in a particular context and with a reader who has the proper background. Generally, reports are rated as more effective when the material is described in clear, basic language (Finn et al., 2001; Harvey, 1997; Ownby, 1990, 1997; Sandy, 1986; Tallent, 1993). Even among readers who have the proper background to understand technical terms, many prefer a more straightforward presentation (Ownby, 1990, 1997; Tallent, 1992, 1993). Technical terms also run the danger of becoming nominalisms in which, by merely naming the phenomenon, persons develop an illusory sense of understanding more than is actually the case. Terms such as *immature* or *sadistic* cover a great deal of information because they are so general, but they say nothing about what the person is like when he or she is behaving in these maladaptive ways. They also do not adequately differentiate one person from the next and are frequently ambiguous. Furthermore, technical terms are often used inappropriately (e.g., a person who is sensitive and cautious in interpersonal relationships is labeled *paranoid,* or *compulsive* is used to describe someone who is merely careful, conscientious, and effective in dealing with details).

W. Klopfer (1960) provides an excellent and still-relevant rationale for using basic English rather than technical terminology. First, and perhaps most important, the use

of basic English allows the examiner, through his or her report, to communicate with and affect a wide audience. This is particularly important because the number and variety of persons who read reports is much greater now than 20 or 30 years ago. Furthermore, basic English is more specific and descriptive of an individual's uniqueness, whereas technical terms tend to deal with generalities. Terms such as *sadomasochistic* and *hostile* do not provide essential information about whether the person is assaultive or suicidal. Finally, the use of basic English generally indicates that the examiner has more in-depth comprehension of the information he or she is dealing with and can communicate this comprehension in a precise, concrete manner. Klopfer stresses that any description found in a psychological report should be comprehensible to any literate person of at least average intelligence. In contrast, psychologists have been found to write in a more technical and complex level when compared with the average client (Harvey, 1997). The first four are examples of translating technical concepts into basic English (Klopfer, 1960):

> "Hostility toward the father figure" becomes "the patient is so fearful and suspicious of people in positions of authority that he automatically assumes an aggressive attitude toward them, being sure that swift retaliations will follow. He doesn't give such people an opportunity to demonstrate their real characteristics since he assumes they are all alike."
>
> "The patient projects extensively" becomes "the patient has a tendency to attribute to other people feelings and ideas originating within himself regardless of how these other people might feel."
>
> "The defenses the patient uses are . . ." becomes "the methods characteristically employed by the patient for reducing anxiety are . . ."
>
> "Empathy" becomes "the patient can understand and sympathize with the feelings of others, since she finds it relatively easy to put herself in their place." (pp. 58–60)
>
> "The client is hostile and resistant" may be changed to include a behavioral description; "when the client entered the room she stated, 'My Dad said I had to come and that's the only reason I'm here' " or "later on in the testing she made several comments such as 'This is a stupid question.' "

The general principle involved in the preceding examples is to translate high-level abstract terms into basic English that provides useful, concrete behavioral descriptions.

Ownby (1990, 1997) recommends combining any conclusion or generalization with specific behaviors or test observations. Recommendations should also be directly linked with the relevant behaviors/generalizations, either in the same place or in the recommendations section. For example, instead of saying a client is "depressed," a writer might state, "The client's behavior, which included self-criticism and occasional crying, suggested he was depressed." Linking generalizations with clear concrete descriptions tends to create reports that are perceived to be relatively credible and persuasive (Ownby, 1990, 1997). If this process is followed, descriptions will be less subject to misinterpretation, less ambiguous, and more likely to convey the unique characteristics of the client. Although abstract technical terms can at times be important components of a psychological report, they should be used sparingly and only when clearly appropriate. This particularly means carefully considering the background of the persons who will read the report. Sandy (1986) even recommends having the clinician collaborate with the relevant recipients of the report, including the client, so that the final report is descriptive rather

than interpretive and the readers are not passive recipients of the "higher" wisdom of the psychologist.

Content Overload

There are no specific rules to follow in determining how much information to include in a report. A general guideline is to estimate how much information a reader can realistically be expected to assimilate. If too many details are given, the information may begin to become poorly defined and vague and, therefore, lack impact or usefulness. When clinicians are confronted with a great variety of data from which to choose, they should not attempt to include it all. A statement such as "The client's relative strengths are in abstract reasoning, general fund of knowledge, short-term memory, attention span, and mathematical computation" is likely to overload the reader with too many details. The clinician should instead adequately develop each of the various points and focus on the areas that are most relevant to the purpose of the report.

Feedback

During the earlier days of psychological assessment, examiners often kept the results of psychological assessments carefully concealed from the client. There was often an underlying belief that the results were too complex and mysterious for the client to adequately understand. In contrast, current practices are to provide the client with clear, direct, and accurate feedback regarding the results of an evaluation (S. Ackerman, Hilsenroth, Baity, & Blagys, 2000; Finn & Tonsager, 1997; Lewak & Hogan, 2003; K. Pope, 1992).

The change toward providing feedback to clients has been motivated by several factors. First, regulations have supported a growing list of consumer rights, including the right to various types of information. Second, it might be perceived as a violation if the client did not receive feedback regarding the results of testing after he or she had been subjected to several hours of assessment. Even the most secure of clients might easily feel uncomfortable knowing a report with highly personal information might be circulated and used by persons in power to make decisions about the client's future. Such practices could understandably result in suspicion and irritation on the part of the public. Third, examiners cannot safely assume that the original referral source will provide feedback to the client. Even if the referral source does provide feedback, there is no guarantee that the information will be provided in an appropriate manner. Thus, the responsibility for providing feedback is ultimately on the clinician. Finally, there is increasing evidence that providing clients with test feedback can result in significant therapeutic benefits (S. Ackerman et al., 2000; Finn & Tonsager, 1992; Gass & Brown, 1992).

The extent to which a clinician providing feedback will allow the client to actually read all or portions of the report varies. The rationale for allowing the client to actually read the report is that doing so enables the client to experience the product of assessment in a direct manner. It also enables a practitioner to explain any areas that are unclear. A significant difficulty is that the client might misinterpret various portions of the report, especially IQ scores and diagnosis. For this reason, most clinicians paraphrase and elaborate on selected portions of the report. This increases the likelihood that clients will

readily understand the most important material and will not be overloaded with too much content.

The likelihood of providing effective feedback can be enhanced by following several guidelines. Initially, the rationale for assessments should be explained and any misconceptions should be corrected. One particularly important misconception is that sometimes clients mistakenly fear that the purpose of assessment is to evaluate their sanity. Practitioners must also select the most essential information to be conveyed to the client. To a large extent, this involves clinical judgment. Important considerations include the client's ego strength, life situation, stability, and receptiveness to different types of material. Typically, three to four general and well-developed areas represent an optimum amount of information. The information that is provided should be carefully integrated into the overall context of the person's life. This integration might be enhanced by providing concrete behavioral examples, reflecting on aspects of the client's behavior, referring to relevant aspects of the client's history, or paraphrasing and expanding on a client's self-descriptions. A useful technique is to have the client evaluate the relevance and accuracy of the information. The client might also be asked to give his or her own examples of the trait or pattern of behavior described in the report. Such collaboration with the client helps the clinician to determine how well the client has understood the feedback. Underlying any feedback should be an attempt to provide the information in a clear, intelligible manner. Commonplace language should be used instead of psychological jargon. It is also important to take into account the client's level of intelligence, education, vocabulary, and level of psychological sophistication. Feedback should be not only a neutral conveyance of data but also a clinical intervention. The information should provide the client with new perspectives and options and should aid in the client's own problem solving.

One possibility is to prepare a personalized report designed specifically for the client. This forces the practitioner to write in a clear, straightforward style. Such reports are more likely to emphasize adaptation rather than pathology. In addition, clear recommendations tend to be emphasized. The optimal communication style is an informal letter written to and for the client ("I am writing to communicate the findings of our psychological assessment . . ."). There are currently available a number of computerized reports directed toward providing the client with feedback. There also seems to be a trend for additional resources to include interpretations directed toward the client such as Lewak, Marks, and Nelson's (1990) *Therapist Guide to the MMPI & MMPI-2: Providing Feedback and Treatments* and Finn's (1996) *Manual for Using the MMPI-2 as a Therapeutic Intervention.* These sources should not be seen as substitutes for a dynamic interaction with a client, but as adjuncts for enhancing this process.

FORMAT FOR A PSYCHOLOGICAL REPORT

Although no single, agreed-on format exists, every report should integrate old information as well as provide a new and unique perspective on the person. Old information should include identifying information (name, birth date, etc.), reason for referral, and relevant history. New information should include assessment results, impressions, summary/conclusions, and recommendations. At the top of the report, practitioners should

indicate its confidential nature by writing "Confidential Psychological Evaluation." A suggested outline follows:

Name:
Age (date of birth):
Sex:
Ethnicity:
Date of Report:
Name of Examiner:
Referred by:
I. Referral Question
II. Evaluation Procedures
III. Behavioral Observations
IV. Background Information (relevant history)
V. Test Results
VI. Impressions and Interpretations
VII. Summary and Recommendations

Although this outline represents a frequently encountered format, there are many variations. Some practitioners prefer to include the client's marital status, occupational status, and handedness (for neuropsychological reports) at the top of the report along with the other demographic information. Other practitioners prefer to exclude the test results section or include additional sections on diagnosis, case formulation, or summary. Still others like to include subheadings in the Impressions and Interpretation section (cognitive functioning and ideation, coping style, affect/mood/emotional control). Sometimes it might be more appropriate to eliminate all or most of the headings and write the report directly to the referral source in a letter format ("Dear Dr. Jones: . . ."). The sample reports included later in this chapter have purposely been chosen to demonstrate a variety of different formats in diverse styles and contexts. Each practitioner needs to develop both the format and style that most effectively meet his or her client's and referral source's needs. In addition, different assessment contexts require different styles and areas of focus.

Referral Question

The Referral Question section provides a brief description of the client and a statement of the general reason for conducting the evaluation. In particular, this should include a brief description of the nature of the problem. If this section is adequately completed, it should give an initial focus to the report by orienting the reader to what follows and to the types of issues that are addressed. This section should begin with a brief, orienting sentence that includes essential information about the client ("Mr. Smith is a 35-year-old, White, married male with a high school education who presents with complaints of depression and anxiety"). Such a sentence clearly and succinctly introduces the client. A prerequisite for this section is that the clinician has developed an

adequate clarification of the referral question. The purpose of testing should be stated in a precise and problem-oriented manner. Thus, phrases such as "the client was referred for a psychological evaluation" or "as a requirement for a class project" are inadequate because they lack focus and precision. It is helpful to include both the specific purpose of the evaluation and the decisions facing the referral source.

Examples of possible reasons for referral include:

- Intellectual evaluation: routine, intellectually disabled (retarded), gifted.
- Differential diagnosis, such as the relative presence of psychological difficulties (i.e., memory problems caused by depression) versus organic impairment (i.e., memory problems because of the early stages of Alzheimer's disease).
- Assessment of the nature and extent of brain damage.
- Evaluation as a component of, and to provide recommendations for, vocational counseling.
- Evaluation of appropriateness for, possible difficulties encountered in, and optimal approach to, psychotherapy.
- Personal insight regarding difficulties with interpersonal relationships.
- Evaluation as an aid in client placement.

These represent general referral questions that, in actual situations, would require further clarification, especially regarding the decisions facing the referral source (see Armengol, 2001). The key should be to find out what the referring person really wants from the report. This may require reading beneath the surface of the referral question(s) and articulating possible hidden agendas and placing the referral question into a wider context than the presenting problem. In some cases, it may be necessary to educate the referral source regarding the strengths as well as the limitations of psychological testing. This may even lead to recommending that the person not be tested. An effective referral question should accurately describe the client's and the referral source's current problems.

After the referral questions have been clarified and outlined, they can be addressed throughout the rest of the report. It is usually helpful to succinctly reiterate and summarize the answers to the referral questions toward the end of the report. Some clinicians prefer to carefully note each of the referral questions that have been made at the beginning of the report, and then bullet/number succinct answers to each of these questions in the Summary and Recommendations. Such a procedure is user friendly, provides succinct answers to the questions, and allows for symmetry and closure to the report.

Evaluation Procedures

The report section that deals with evaluation procedures simply lists the tests and other evaluation procedures used but does not include the actual test results. Usually, full test names are included along with their abbreviations. Later in the report, the abbreviations can be used, but the initial inclusion of the entire name provides a reference for readers who may not be familiar with test abbreviations. For legal evaluations or other occasions in which precise details of administration are essential, it is important to include

the date on which different tests were administered and the length of time required to complete each one. For most routine evaluations, however, this degree of detail is not recommended. It may also be important to include whether a clinical interview or mental status examination was given and, if so, the degree of interview structure and the amount of time required for the interview or examination. Evaluation procedures may not necessarily be restricted to testing and interviews with the client. Often, evaluation includes a review of relevant records such as medical reports, nursing notes, military records, police records, previous psychological or psychiatric reports, or educational records. Additional material might come from interviews with individuals such as spouses, children, parents, friends, employers, physicians, lawyers, social workers, or teachers. If any of these sources are used, their dates and, if relevant, who wrote them should be included. This section might end with a statement summarizing the total time required for the evaluation.

Behavioral Observations

A description of the client's behaviors can provide insight into his or her problem and may be a significant source of data to confirm, modify, or question the test-related interpretations. These observations can be related to a client's appearance, general behavioral observations, or examiner-client interaction. Descriptions should be tied to specific behaviors and should not represent a clinician's inferences. For example, instead of making the inference that the client was "depressed," it is preferable to state that "her speech was slow and she frequently made self-critical statements such as 'I knew I couldn't get that one right.' "

Relevant behavioral observations made during the interview include physical appearance, behavior toward the task and examiner, and degree of cooperativeness. A description of the client's physical appearance should focus on any unusual features relating to facial expressions, clothes, body type, mannerisms, and movements. It is especially important to note any contradictions, such as a 14-year-old boy who acts more like an 18-year-old or a person who appears dirty and disheveled but has an excellent vocabulary and high level of verbal fluency. The behaviors the client expresses toward the test material and the examiner often provide a significant source of information. These may include behaviors that reflect the person's level of affect, manifest anxiety, presence of depression, or degree of hostility. The client's role may be as an active participant or generally passive and submissive; he or she may be very much concerned with his or her performance or relatively indifferent. The client's method of problem solving is often a crucial area to note, and it may range from careful and methodical to impulsive and disorganized. It is also important to pay attention to any unusual verbalizations that the client makes about the test material. The level of cooperation expressed by the client should be a factor in assessing the validity of the test results. This is especially important for intelligence and ability tests, because a prerequisite is that the client be alert and attentive, and put forth his or her best effort. It may also be important to note events before testing, such as situational crises, previous night's sleep, or use of medication. If there are situational factors that may modify or bring into question the test's validity, they should be noted with statements such as, "The test results should be viewed with caution because . . ." or "The degree of maladjustment indicated on the test scores may

represent an exaggeration of the client's usual level of functioning due to conditions surrounding the test administration." Often, the most important way to determine test validity in relationship to the client is through a careful look at the client's behaviors relating to the tests and his or her life situation before testing.

Sattler (1992) has developed a Behavior and Attitude Checklist comprising 10 major categories that can be rated on a seven-point scale (see Table 15.2). The examiner may wish to use this checklist as a tool to help focus on areas that might be significant to mention or discuss. It is important to emphasize that other crucial behaviors that are not covered in the checklist may occur and still require discussion.

Behavioral observations should usually be kept concise, specific, and relevant. If a description does not allow for some insight about the person or demonstrate his or her uniqueness, it should not be included. Thus, if a behavior is normal or average, it is usually not important to discuss other than to briefly mention that the person had, for example, an average level of cooperation, alertness, or anxiety. The focus, then, should be on those client behaviors that create a unique impression. The relative length of this section varies from a few brief sentences to considerably longer depending on the amount of relevant information the clinician has noticed. The relative importance of this section in relationship to the overall report is, likewise, extremely varied. Sometimes, this section can be almost as important as the test results, whereas at other times, it might consist of a few minor observations.

Clinicians who prefer behavioral assessment procedures might wish to emphasize the behavioral observation section by providing more in-depth descriptions of relevant antecedents. In addition, consequent events surrounding the problem behavior itself might be evaluated in relationship to their onset, duration, frequency, and intensity. Specific strategies of behavioral assessment include narrative descriptions, interval recording, event recording, ratings recordings, and self-report inventories (see Chapter 4).

Some examiners may wish to summarize information from a Mental Status Examination in the Behavioral Observations section. In these cases, there is necessarily a movement away from concrete descriptions of behaviors to inferences about these behaviors. For example, a clinician may infer, based on behavioral observations, that the client was oriented to time and place. Additional categories might include verbalizations, psychomotor activity, affect, thought processes/contents, and insight/judgment (see section on the Mental Status Examination in Chapter 3).

Another exception to adhering exclusively to concrete behavioral descriptions is that, at the end of the Behavioral Observations section, it is customary and appropriate to include a statement indicating the validity of the assessment procedures. For example, it might state something like: "Given the consistency and detail of the client's responses, the client's high level of motivation, and validity indicators on the MMPI-2, the assessment appears to be an accurate assessment of this person's current level of functioning."

Background Information (also Referred to as Relevant History)

The write-up of a client's background information should include aspects of the person's history that are relevant to the problem the person is confronting and to the interpretation of the test results. The history, along with the referral question, should also

Table 15.2 Behavior and Attitude Checklist

Client's name: _____ Examiner: _____

Age: _____ Date of report: _____

Test(s) administered: _____ Date of examination: _____

IQ: _____ Grade: _____

Instructions: Place an X on the appropriate line for each scale.

I.	*Attitude toward examiner and test situation:*		
	1. cooperative	____ : ____ : ____ : ____ : ____ : ____ : ____	uncooperative
	2. passive	____ : ____ : ____ : ____ : ____ : ____ : ____	aggressive
	3. tense	____ : ____ : ____ : ____ : ____ : ____ : ____	relaxed
	4. gives up easily	____ : ____ : ____ : ____ : ____ : ____ : ____	does not give up easily
II.	*Attitude toward self:*		
	5. confident	____ : ____ : ____ : ____ : ____ : ____ : ____	non confident
	6. critical of own work	____ : ____ : ____ : ____ : ____ : ____ : ____	accepting of own work
III.	*Work habits:*		
	7. fast	____ : ____ : ____ : ____ : ____ : ____ : ____	slow
	8. deliberate	____ : ____ : ____ : ____ : ____ : ____ : ____	impulsive
	9. thinks aloud	____ : ____ : ____ : ____ : ____ : ____ : ____	thinks silently
	10. careless	____ : ____ : ____ : ____ : ____ : ____ : ____	neat
IV.	*Behavior:*		
	11. calm	____ : ____ : ____ : ____ : ____ : ____ : ____	hyperactive
V.	*Reaction to failure:*		
	12. aware of failure	____ : ____ : ____ : ____ : ____ : ____ : ____	unaware of failure
	13. works harder after failure	____ : ____ : ____ : ____ : ____ : ____ : ____	gives up easily after failure
	14. calm after failure	____ : ____ : ____ : ____ : ____ : ____ : ____	
	15. apologetic after failure	____ : ____ : ____ : ____ : ____ : ____ : ____	not apologetic after failure
VI.	*Reaction to praise:*		
	16. accepts praise gracefully	____ : ____ : ____ : ____ : ____ : ____ : ____	accepts praise awkwardly
	17. works harder after praise	____ : ____ : ____ : ____ : ____ : ____ : ____	retreats after praise
VII.	*Speech and language:*		
	18. speech poor	____ : ____ : ____ : ____ : ____ : ____ : ____	speech good
	19. articulate language	____ : ____ : ____ : ____ : ____ : ____ : ____	inarticulate language
	20. responses direct	____ : ____ : ____ : ____ : ____ : ____ : ____	responses vague
	21. converses spontaneously	____ : ____ : ____ : ____ : ____ : ____ : ____	only speaks when spoken to
	22. bizarre language	____ : ____ : ____ : ____ : ____ : ____ : ____	reality-oriented language
VIII.	*Visual-motor:*		
	23. reaction time slow	____ : ____ : ____ : ____ : ____ : ____ : ____	reaction time fast
	24. trial-and-error	____ : ____ : ____ : ____ : ____ : ____ : ____	careful and systematic
	25. skillful movements	____ : ____ : ____ : ____ : ____ : ____ : ____	awkward movements
IX.	*Motor:*		
	26. defective motor coordination	____ : ____ : ____ : ____ : ____ : ____ : ____	good motor coordination
X.	*Overall test results:*		
	27. reliable	____ : ____ : ____ : ____ : ____ : ____ : ____	unreliable
	28. valid	____ : ____ : ____ : ____ : ____ : ____ : ____	invalid

Source: From *Assessment of Children* (3rd ed.), p. 92, by J. M. Sattler, 1992, San Diego: Author. Copyright 1992 by Jerome M. Sattler, Publisher. Reprinted by permission.

place the problem and the test results into the proper context. In accomplishing these goals, the clinician does not need to include a long, involved chronology with a large number of details, but rather should be as succinct as possible. Some practitioners even urge that the background information section be kept to one concise paragraph, particularly in medical settings where there is considerable emphasis on conciseness. In selecting which areas to include and which to exclude, a clinician must continually evaluate these areas in relationship to the overall purpose of the report. It is difficult to specify precise rules because each individual is different. Furthermore, each clinician's own personal and theoretical orientation alters the types of information he or she feels are significant. Whereas one clinician may primarily describe interpersonal relationships, another may focus on intrapsychic variables, birth order, early childhood events, or details about the client's present situation and environment. The key is to maintain a flexible orientation so that the interviewer is aware of the most significant elements in the client's life. In general, the end product should include a good history of the problem, along with areas such as important life events, family dynamics, work history, personal interests, daily activities, and past and present interpersonal relationships (see Table 3.1 in Chapter 3).

When describing a client's background, it is important to specify where the information came from ("The client reported that . . ."). This is particularly essential when there may be some question regarding the truth of the client's self-reports or when the history has been obtained from multiple sources.

Usually, a history begins with a brief summary of the client's general background. This can be followed by sections describing family background, personal history, medical history, history of the problem, and current life situation.

The extent to which a clinician decides to pursue and discuss a client's family background is subject to a great degree of variability. The primary purpose of such information is to help determine causal factors, what variables might help maintain relevant behaviors, and the extent to which the family should be used as either a focus of systemic intervention or as social support. At a minimum, a brief description of the client's parents is warranted; this may include whether they are separated/divorced and alive/deceased, and their socioeconomic level, occupation, cultural background, and health status. Sometimes, it is important to include information about the emotional and medical backgrounds of parents and close relatives, because certain disorders occur with greater frequency in some families than in the overall population. A description of the general atmosphere of the family is often helpful, including the client's characteristic feelings toward family members and his or her perceptions of their relationships with each other. Descriptions of common family activities and whether the family lived in an urban or a rural environment might also be included. If one or both parents died while the client was young, the clinician can still discuss the speculations the client has about his or her parent(s) and can describe the significant persons for the client as he or she was growing up.

The client's personal history can include information from infancy, early childhood, adolescence, and adulthood. Each stage has typical areas to investigate and problems to be aware. The information from infancy usually either represents vague recollections or is secondhand information derived from parents or relatives. Thus, it may be subject to a great deal of exaggeration and fabrication. If possible, it may be helpful to have

details verified by additional sources, such as through direct questioning of parents or examination of medical records. The degree of contact with parents, family atmosphere, and developmental milestones may all be important areas to discuss. Because physical and psychological difficulties often are related and occur simultaneously, including a client's early medical history is sometimes helpful. The most significant tasks during childhood are the development of peer relationships and adjustment to school. What was the quality of the client's early friendships? How much time did the client spend with others? Were there any fights or rebellious acting out? Was the client basically a loner, or did he or she have a large number of friends? Did the person join clubs and have group activities, hobbies, or extracurricular interests? In the academic area, it may be of interest to note the usual grades, best or worst subjects, and whether the client skipped or repeated grades. Furthermore, what was his or her relationship with parents, and did the parents restrict activities or allow relative freedom? During adolescent years, clients typically face further academic, psychological, and social adjustments to high school. Of particular importance are their reactions to puberty and early heterosexual relationships. Did they have difficulties with sex role identity, abuse drugs or alcohol, or rebel against authority figures? The adult years center around occupational adjustment and establishing marital and family relationships. During early adulthood, what were clients' feelings and aspirations regarding marriage? What were their career goals? Did they effectively establish independence from parents? As adulthood progressed, were there any significant changes in the quality of their close relationships, employment, or expression of sexuality? What activities did they engage in during their leisure time? As clients age, they face challenges in adapting to their declining abilities and limitations, and developing a meaningful view of their lives.

Although the personal history can help place the problem in its proper context and explain certain causative factors, it is usually essential to spend some time focusing directly on the problem itself. Of particular importance are the initial onset and the nature of the symptoms. From the time the client first noticed these symptoms, have there been any changes in frequency, intensity, or expression? If a formal diagnosis will be made, it is particularly important to have a clear description of symptom patterns to substantiate such a diagnosis. It might also be important to determine whether there were any previous attempts at treatment, and if so, what was the outcome? In some reports, the history of the problem is the longest and most important part of the history section.

The family and personal histories usually reveal information relating to the predisposing cause of a client's difficulties, whereas the history of the problem often provides an elaboration of the precipitating and reinforcing causes. To complete this picture, the clinician also has to develop a sense of the factors currently reinforcing the problem. This requires information relating to the client's life situation. Significant areas may be the client's life stresses, including changes that he or she is confronting. In addition, what are the nature of and resources provided by his or her family and work relationships? Finally, it is important to understand the alternatives and decisions that the client is facing.

Sometimes, an evaluation needs to assess the possible presence and nature of organic impairment. In many cases, the history is of even greater significance than test results; and, often, the most valuable information a psychologist can provide to a referring medical practitioner is a thorough history. Thus, the history needs to be complete

and must address a number of areas that are not ordinarily covered in personality evaluations. Several interview aids have been commercially developed to help ensure that most relevant areas are covered (see Chapter 12). If the person reports having had a head injury, it is important to note the length of time the client was unconscious (if at all), whether he or she actually remembers getting hit, the last memory before the injury, and the first thing he or she clearly remembers following the injury. In all neuropsychological assessments, a crucial area is to establish the person's premorbid level of functioning. This may mean obtaining information on his or her grade point average in high school or college, sending for any relevant records (e.g., previous IQ results), and determining previous highest level of employment and personal interests or hobbies. Often, it may be necessary to verify the client's previous level of functioning from outside sources, such as from parents or employers. In determining the probable cause of brain impairment, it may be difficult to rule out other possibilities, such as exposure to toxic substances, strokes, high fevers, or other episodes of head trauma. Areas of current functioning that need to be addressed might include memory problems, word-finding difficulties, weakness on one side of the body, alterations in gait, loss of consciousness, and unusual sensations. Previous assessments with CT/MRI scans, EEGs, or neurological physical exams would also be important to obtain. Even though these medical records might be able to identify the site and size of a lesion, it is still the work of the psychologist to describe what the person is doing as a result of these lesions. It might also be important to obtain current or past information regarding drug intake, especially recent alterations in prescriptions, because these might affect psychological functioning. The interview data and neuropsychological test results from a psychologist should ideally be combined with and complement medical records, such as CT scans and neurological exams. Although the preceding topics are by no means exhaustive, they represent some of the more important areas to consider when taking a history related to possible neuropsychological deficit.

Although the quantity of such information may seem immense, the history format described here is only a general guideline. At times, it may be appropriate to ignore many of the areas mentioned earlier and focus on others. In condensing the client's history into the report, it is important to avoid superfluous material and continually question whether the information obtained is relevant to the general purpose of the report. Again, some practitioners might prefer to restrict the length of this section to a single paragraph. Furthermore, it is typically not useful to include material that is already in the possession of the referral source. This may be perceived as needless duplication and could result in the history section's becoming needlessly long.

Test Results

For certain reports, it may not be necessary to list test scores. Some practitioners even prefer to completely exclude actually giving test scores because it might give the impression that the report is too data/test oriented. However, it is usually recommended that, at some point, test scores be included, especially in legal reports or when professionals who are knowledgeable about testing will read the report. If practitioners feel that including this section in the report itself is too test/data oriented or serves to unnecessarily "clutter" the report, the test data, perhaps including the actual profiles,

might be inserted in an appendix. This could then be noted and the reader referred to the appendix in the evaluation procedures section.

If actual test scores are included, standard (rather than raw) scores should be the mode of presentation. Referral sources have consistently indicated that percentiles are preferred over other types of standard scores (Finn et al., 2001). Because various tests use somewhat different types of standard scores, it is recommended that each set of test scores include both the standard score and percentiles. Clinicians may also wish to indicate the relative magnitude of the relevant scores ("Very high," "High," etc.) or whether the scores exceed some clinically meaningful cutoff.

Intelligence test scores are traditionally listed first and, for the Wechsler scales, should include IQ scores (Verbal, Performance, Full Scale IQ), index scores, and subtest scaled scores. Subtests that have been found to be significant strengths should be indicated with an "S" next to the subtest score and significant weaknesses should be indicated with a "W." This is often followed by other cognitive test results such as the Bender or Wechsler Memory Scale-III. Bender results can simply be summarized by a statement such as: "Empirically not in the organic range, although there were difficulties organizing the designs and frequent erasures." MMPI-2/MMPI-A results are often listed in the order in which they appear on the profile sheet. Objective personality tests (MMPI-2, MMPI-A, CPI) should always be referred to by their standardized (usually *T*) scores and not their raw scores. Whereas it is fairly straightforward to list the objective and intelligence test scores, it is considerably more difficult to adequately describe the scores on projective tests. The Rorschach summary sheet can be included, but the results from projective drawings and the TAT are usually omitted. Should a clinician wish to summarize projective drawings, a brief statement is usually sufficient, such as "Human figure drawings were miniaturized and immature, with the inclusion of two transparencies." Likewise, TAT "scores" can be summarized by a brief statement of the strongest needs and press, and a mention of the most common themes encountered in the stories.

Impressions and Interpretations (also Referred to as Discussion)

This section can be considered the main body of the report. It requires that the main findings of the evaluation be presented in the form of integrated hypotheses. The areas discussed and the style of presentation vary according to the personal orientation of the clinician, the purpose of testing, the individual being tested, and the types of tests administered. As emphasized previously, assessment data should be organized according to different integrated topics or presented as a chronological narrative of the person. In contrast, a test-by-test presentation is strongly discouraged. To organize the information from an assessment, W. Klopfer (1960) recommends using a grid with the topics for consideration in the left column (derived from Table 15.1) with the assessment results in the top row. This enables the practitioner to extract essential findings from the data and list them in the appropriate box where the topic and the method of assessment intersect. When actually writing the Impressions and Interpretations section of the report, the clinician can then review all findings in a particular topic and summarize them on the report. An example of such a grid is given in Table 15.3. The list of assessment

Table 15.3 Sample grid of assessment domains by tests administered

	Evaluation Procedures				
Topics	Interview	WAIS-III	MMPI-2	BDI	Rorschach
Validity of results					
Cognitive functioning					
Emotional controls					
Interpersonal relations					
Diagnostic impression					
Recommendations					

Source: Adapted from Klopfer (1960), *The Psychological Report,* New York: Grune & Stratton.

methods is dependent on which tests the examiner administered, but the topics can be chosen and arranged according to areas the clinician would like to focus on.

All inferences made in the Impressions and Interpretations section should be based on an integration of the test data, behavioral observations, relevant history, and additional available data. The conclusions and discussion may relate to areas such as the client's overt behavior, self-concept, family background, intellectual abilities, emotional difficulties, medical disorders, school problems, or interpersonal conflicts. A client's intellectual abilities often provide a general frame of reference for a variety of personality variables. For this reason, a discussion of the client's intellectual abilities usually occurs first. Although this should include a general estimate of the person's intelligence as indicated by IQ scores, it is also important to provide a discussion of more specific abilities. This discussion may include an analysis of areas such as memory, problem solving, abstract reasoning, concentration, and fund of information. If the report will be read by persons who are familiar with test theory, it may be sufficient to include IQ scores without an explanation of their normative significance. In most reports, it is helpful to include the IQ scores as well as the percentile ranking (see Appendix B) and general intellectual classification (high average, superior, etc.; see Table 5.1). Some examiners may even prefer to omit the actual IQ scores in favor of including only percentile rank and general classification. This can be useful in cases in which persons reading the report might be likely to misunderstand or misinterpret unexplained IQ scores. After a general estimate of intelligence has been made, it should, whenever possible, be followed by a discussion of the client's intellectual strengths and weaknesses. This may involve elaborating on the meaning of the difference between Verbal IQ and Performance IQ or a discussion of subtest scatter. In addition, it can be useful to compare the client's potential level of functioning with his or her actual performance. If there is a wide discrepancy between these two, reasons for this discrepancy should be offered. For example, the client may be underachieving because of anxiety, low motivation, emotional interference, or perceptual processing difficulties. Practitioners may also wish to discuss additional noncognitive areas of intellectual assessment. This might include the extent to which the person prefers to achieve through independent activities versus a structured environment, the level of motivation, or, the relative intellectual efficiency or hardiness. Cognitive assessments in psychiatric contexts might include any bizarre associations,

degree to which the person's thinking is organized, or how concrete or abstract his or her thought processes are.

Whereas a discussion of intellectual abilities is relatively clear and straightforward, the next sections are frequently more difficult to select. There are an extremely wide number of possibilities to choose from, many of which are listed in Table 15.1. Some practitioners recommend including set topics. These typically include the client's level of cognitive functioning, emotional functioning (affect and mood), and interpersonal relationships. A neuropsychological evaluation might divide the impression and interpretation into areas such as memory, language functions, executive abilities, awareness of deficits, sensory/perceptual functions, and personality (Groth-Marnat, 2000a; Hebben & Milberg, 2002). One rationale for not having a preset list of topics to discuss is that the topics should be based primarily on the referral question. This allows the practitioner to flexibly organize the topics based on the context of the report and the needs of the referral source and client. If the referral question is clearly focused on a specific problem, it may be necessary to elaborate on only two or three topics. A referral question that is more general may require a wider approach in which six or more areas are discussed.

Some additional common and important topics are the client's level of psychopathology, dependency, hostility, sexuality, interpersonal relationships, diagnosis, and behavioral predictions. A client's level of psychopathology refers to the relative severity of the disturbances he or she is experiencing. It is important to distinguish whether the results are characteristic of normals, outpatients, or inpatients, and whether the difficulties are long term or a reaction to current life stresses. Does the client use behaviors that are adaptive or those that are maladaptive and self-defeating? Within the area of ideation, are there persistent thoughts, delusions, hallucinations, loose associations, blocking of ideas, perseveration, or illogical thoughts? It may also be important to assess the adequacy of the client's judgments and relative degree of insight. Can the person effectively make plans, understand the impact he or she has on others, and judge the appropriateness of his or her behavior? To assess the likelihood of successful therapy, it is especially important to assess the client's level of insight. This includes assessing the person's ability to think psychologically, awareness of his or her own changing feelings, understanding of the behaviors of others, and ability to conceptualize and discuss relevant insights.

Usually, a client's greatest conflicts center on difficulties with dependency, hostility, and sexuality. In discussing a client's dependency, it is important to discuss the strength of these needs, the typical roles played with others, and present or past significant relationships. In what ways does the client defend himself or herself against, or cope with, feelings of dependency? This evaluation may include a discussion of defense mechanisms, thoughts, behaviors, feelings, or somatic responses as they relate to dependency. The relative intensity of a client's hostility is also important. Is the expression of hostility indirect, or is it direct in the form of either verbal criticisms or actual assaultive behavior? If the expression of hostility is covert, it may be the result of factors such as fear of loss of love, retaliation, or guilt. When the client does feel anger, what are his or her characteristic defenses against these feelings? For example, some clients might express opposite behaviors, with overly exaggerated concern for others, or they might direct their anger inward by developing physical aches and pains that serve as self-punishment for having aggressive impulses. They may also adapt through means such as extreme suspiciousness of others, which has been created by denying their hostility and attributing it to others. A discussion of a client's sexuality usually involves

noting the relative intensity of his or her urges and the degree of anxiety associated with the expression of those urges. Does the client inhibit his or her sexuality because of a belief that it is dirty, experience anxiety over possible consequences, or associate it with aggressiveness? Defenses against sexual urges may be handled in ways similar to hostility, such as by performing the opposite behavior through extreme religiosity or celibacy, or by denying the feelings and attributing them instead to others. On the other hand, clients may impulsively act out their sexual urges, at least in part, out of a need to obtain self-affirmation through sexual contact. Clinicians may want to discuss the dynamics involved in any unusual sexual practices.

Discussing clients' characteristic patterns and roles in interpersonal relationships can also be extremely useful. These can often be discussed in relationship to the dimensions of submissiveness/dominance and love/hate, or the extent to which they orient themselves around the need to be included, control others, or seek affection. Is their style of communicating typically guarded, or is it open and self-disclosing to the extent that they can discuss areas such as painful feelings and fears? Can they deal with the specifics of a situation, or are they usually vague and general? Do they usually appear assertive and direct, or passive and indirect? Finally, it is often important to determine the extent to which they are perceptive about interpersonal relationships and their typical approaches toward resolving conflict.

It may also be appropriate to include descriptions of vocational goals and aptitudes. This is becoming increasingly important in educational reports, especially for students with special educational needs, such as those with disabilities (Schalock et al., 1994). Many of the tests covered in this text can help in assessing a person's strengths and weaknesses, but practitioners may also need to include further assessment devices, such as the Self-Directed Search, Strong-Campbell Interest Inventory, or Kuder Occupational Interest Survey (see Prince & Heisser, 2000).

A frequent consideration is whether the client's difficulties will continue or, if currently absent, recur. If the client's future prospects are poor, a statement of the rationale for this conclusion should be given. For example, if a clinician predicts that the response to treatment will be poor, he or she should explain that this is caused by factors such as a strong need to appear hypernormal, poor insight, and a high level of defensiveness. Likewise, favorable predictions should include a summary of the client's assets and resources, such as psychological mindedness, motivation to change, and social supports. If difficulties are likely to be encountered during the course of treatment, the nature and intensity of these difficulties should be discussed. The prediction of suicidal potential, assaultive behavior, child abuse, or criminal behavior is essential in certain types of reports. Often the tests themselves are not useful in predicting these behaviors. For example, one of the best ways of predicting suicidal potential is to evaluate the client's history, current environment, personal resources, and degree of suicide intent (Klespies & Dettmer, 2000; Stelmachers, 1995). However, research indicates that many predictions of behavior, such as dangerousness, are subject to error (Binder, 1999; Freedman, 2001; Megargee, 1995). This is especially true for long-term predictions. Clinicians should thus exercise appropriate caution in making predictions and not exceed the bounds of reasonable certainty.

Sometimes clinicians may wish to include a separate section on diagnosis. However, whether to include a *DSM-IV* (1994; or *ICD-10;* World Health Organization, 1990) diagnosis has been an area of some controversy. Some clinicians feel that labels should be

avoided because they may result in self-fulfilling prophecies, be overly reductionistic, and allow clients to avoid responsibility for their own behavior. Other objections to diagnosis stem from researchers who feel that many of the terms are not scientifically valid (Beutler & Malik, 2002; Rosenhan, 1973) and are not particularly useful in planning interventions (Beutler & Malik, 2002; Groth-Marnat, Roberts, & Beutler, 2001; Houts, 2002). If a clinician does decide to give a diagnosis, he or she must first have a clear operational knowledge of the diagnostic terms. He or she should also include the client's premorbid level of adjustment, and the severity and frequency of the disturbance. Instruments such as the Structured Clinical Interview for the *DSM-IV* (SCID; First, Spitzer, et al., 1996; First et al., 1997), Structured Interview for *DSM-IV* Personality (SIDP), or Anxiety Disorders Interview Schedule (T. Brown et al., 1994) might help to increase the reliability of diagnosis. It is also important to include the possible causes of the disorder. A discussion of causes should not be simplistic and one-dimensional but rather should appreciate the complexity of causative factors. Thus, causes may be described from the perspective of primary, predisposing, precipitating, and reinforcing factors. Clinicians may also discuss the relative significance of biological, psychological, and sociocultural variables.

Summary and Recommendations

The purpose of the summary subsection is to restate succinctly the primary findings and conclusions. This requires that the practitioner select only the most important issues and that he or she be careful not to overwhelm the reader with needless details. As emphasized previously, a useful strategy in the summary section is to provide brief bulleted/numbered answers to each of the referral questions. It should also be noted that some practitioners prefer to place the summary subsection at the end of the Impressions and Interpretations section. Either location is acceptable; the choice can be based on the clinician's personal preference and the needs of the report as suggested by the referral question(s) and background of the readers.

The ultimate practical purpose of the report is contained in the recommendations because they suggest what steps can be taken to solve problems. Such recommendations should be clear, practical, and obtainable, and should relate directly to the purpose of the report. The best reports are those that help the referral sources and/or the clients solve the problems they are facing (Armengol, Moes, et al., 2001; Finn et al., 2001; Ownby, 1997; Tallent, 1993). To achieve this report-writing goal, the clinician must clearly understand the problem, the best alternatives for remediation, and the resources available in the community. One practical implication is that writers can improve their reports by becoming as familiar as possible with the uses to which their reports will be applied. An effective report must answer the referral question and have decisional value. After these factors have been carefully considered, recommendations can be developed. Decisions related to recommendations occur on three different levels (Beutler, 1995). First, decisions need to be made related to the *setting or context* (outpatient, day hospital, halfway house, inpatient, new work environment, change in schools/classes). Second, consideration needs to be given to *developing a relationship* with the client (degree of resistance, level of insight, interpersonal style, empathy, etc.). Finally, decisions need to be made about *specific intervention procedures* (systematic desensitization, emotional support, vocational training, rehabilitation, special education, etc.).

Practitioners and researchers alike have become increasingly interested in tailoring the results of assessment toward optimal client interventions (see Chapter 14). For example, Beutler et al. (2000) provide strong empirical support that planning interventions around problem complexity, problem severity, motivational distress, coping styles, level of resistance (reactance), and social support can increase the effectiveness of treatment outcomes. Additional relevant variables include the stage of change (Prochaska & DiClemente, 1992), hypnotic responsiveness (Groth-Marnat, 1991), neuroticism, level of stress (Wickramasekera, 1995a), level of abstract versus concrete thinking, and ego strength. Outcome research for particular types of problems (anxiety, sexual dysfunction, etc.) frequently indicates that certain interventions are more effective than others. Especially relevant are the specific causes and expression of a disorder. For example, depression can result from and be maintained by a variety of factors including irrational cognitions, poor social skills, grief, stress, poor interpersonal relations, medical illnesses, substance abuse, medications, and faulty biological mechanisms. The relative contribution of each of these causes has important implications for optimally tailoring client interventions. By considering each of these areas, the practitioner can develop optimal specific, targeted recommendations.

One clear finding is that reports are typically rated most useful when their recommendations are highly specific rather than general (Armengol, 2001; Finn et al., 2001; Ownby, 1990, 1997; G. White, Nielsen, & Prus, 1984). Thus, a recommendation that states "The client should begin psychotherapy" is not as useful as a statement of the need for "individual therapy focusing on the following areas: increased assertiveness, relaxation techniques for reducing anxiety, and increased awareness of the self-defeating patterns he creates in relationships." Likewise, a recommendation for "special education" can be improved by expanding it to "special education two hours a day, emphasizing exercises in auditory sequencing and increasing immediate recall for verbally relevant information." However, caution should be exercised when providing specific recommendations in some contexts because some health professionals may feel that developing treatment recommendations is primarily their responsibility or perhaps should be made by the overall treatment team (Tallent, 1993). After the report, with its recommendations, has been submitted, continued contact should be made with the readers(s) to make sure the report has not been filed and forgotten. Even the best report is not functional unless the recommendations are practical, obtainable, and actually put into action.

SAMPLE REPORTS

The sample reports in this section are from the more common settings in which clinicians work and consult. The dimensions in which the reports vary are:

- Format.
- Referral question.
- Extent to which history rather than test data is emphasized.
- Types of tests used.
- Degree to which they include a variety of descriptions rather than being case-focused with a relatively limited range of topics.

In each setting, specific questions have been presented, along with decisions that must be made related to the client. The different reports illustrate how the clinician has integrated the test data, client's history, and behavioral observations to handle these questions. The reports were selected to illustrate a wide diversity in format, length, type of setting, referral question, and type of tests used.

The first report was developed for a psychiatric setting and was intended to be read by professional mental health personnel. For this reason, there is some use of technical language and a focus on developing a detailed, traditional *DSM-IV* diagnosis. What is noteworthy in the handling of the test data is that the bulk of the discussion relating to test interpretation revolves around projective test findings (Rorschach, TAT, projective drawing). Furthermore, much of the projective data is used in a qualitative, content-oriented manner. This was achieved by providing actual verbatim responses, which give a more colorful and rich portrayal of the client's thought processes than could be achieved through quantitative scores. For example, some of the TAT stories are written out to illustrate attitudes toward the client's parents and how the client perceives and attempts to cope with his inner sense of "evilness." This is a more test-oriented approach than is recommended, but the report has the advantage of allowing the readers to see why and how certain inferences were made. For example, relevant Rorschach indices and TAT stories have been included adjacent to interpretations based on this data. For this reason, it is an excellent teaching case.

The second report, written in a legal context, critiques a report done by another professional who had evaluated a client for possible malingering. This is an important teaching case because, in the process of evaluating the report done by another professional, there is information on how the assessment should have optimally been performed. As such, it provides guidelines for proper assessment procedures. Specific points are the importance of using norms similar to the client that is being evaluated, the use of converging sources of information, placing the results into a wider context, and following acknowledged protocols for assessing client domains (in this case, malingering). The professional in question had used inappropriate norms (norms were for brain-injured patients complaining of memory difficulties and not pain patients), had used only one source of information (only one test of malingering), and had not considered the wider context (client characteristics, medical records). One feature to pay particular attention to is the clear listing of referral questions at the beginning of the report. This is followed in the body of the report by numbered paragraphs that respond to each of the referral questions. The main points are then summarized in bulleted form at the end of the report. Another aspect of the report is its format—written as an informal letter rather than according to structured headings (i.e., referral question, evaluation procedures, etc.).

The third sample is from an educational context; the client (a 12-year-old Hispanic female) was experiencing emotional and behavioral problems that were impacting her academic performance and peer relations. This report is an important addition because it not only includes a child, but also demonstrates how the examiner handled ethnic diversity. There is also information on how the combination of emotional difficulties and being bilingual was likely to have lowered her optimal level of cognitive performance. Two noteworthy strengths of the report were the extent to which a wide variety of sources of information were used (previous records, interviews with

several different persons, a number of formal tests) and the clear description of the client's assets. The formal tests were targeted toward specific areas (intelligence, achievement, trauma, self-perception) and used a combination of objective and projective/qualitative information.

The final report was written in a general psychology clinic and was intended for use by mental health professionals. As a result, some technical language is used, primarily in the form of a formal *DSM-IV* diagnosis. The major feature of the report is the extensive development of a detailed treatment plan for psychotherapeutic intervention. This plan was developed based on the Systematic Selection Model detailed in Chapter 14. The recommendations are eclectic in orientation and assume that the treating practitioner can effectively use a number of techniques from a variety of theoretical orientations. Another feature of the report is the absence of psychological test data. Specifically, there is no Test Results section because some, if not many, practitioners believe that the inclusion of detailed test results is both unnecessary and results in cluttering up the report with distracting detail. It is rather assumed that the referral source is most interested in the integration of the overall assessment along with the relevant recommendations.

THE PSYCHIATRIC SETTING*

NAME: Robert
DATE OF BIRTH: 2/5/86
DATES OF EVALUATION: 3/14/2000

REFERRAL QUESTIONS
Robert M. is a 14-year-old high school student admitted to the psychiatric unit at Monte Hospital on March 10, 2000, for his second psychiatric hospitalization. He was referred for an interview and psychological testing by Harold Smith, MD, to estimate his intelligence, differential diagnosis, behavioral dynamics, and potential for adjustment.

EVALUATION PROCEDURES
Robert was interviewed for one hour and administered the Wechsler Intelligence Scale for Children-3rd ed. (WISC-III), Rorschach, Thematic Apperception Test (TAT), Bender Visual Motor Gestalt Test (Bender) and Draw-A-Family Test. Information was also derived from a review of his hospital charts and a conversation with his psychiatrist (15 minutes; March 9, 2001).

BEHAVIORAL OBSERVATIONS
Robert was seen on the Delta unit at Monte Hospital on March 12, 2000. He presented himself looking his stated age and adequately nourished as a pubescent male with somewhat disheveled dark blond hair parted approximately on the right side, a fair facial complexion with mild acne and dressed in a white T-shirt, a blue and white plaid long-sleeved shirt in poor repair and new-appearing Levi's. A cross-shaped earring

* Report contributed by Tom MacSpeiden, PhD.

dangled from his left ear. At the beginning of the interview and multiple times thereafter, Robert asked when the evaluation would end. At one point, this examiner asked what Robert would do after the evaluation was completed, and Robert said he would go to his room to lie down because he was tired. He had a depressed facial expression, but was physically agitated and demonstrated considerable motor overflow although he remained continuously seated in his chair. During the intelligence testing, he seemed to exert effort until a task became difficult whereupon he became disheartened and reduced his effort. During the projective testing, he was brief and hurried to complete the tasks. There were no indications of delusions or hallucinations, and he was oriented in all three spheres. His hygiene was adequate. He was alert, and both recent and remote memory appeared intact. Given the previous behavioral observations and additional test indicators, I believe the assessment is an accurate assessment of his current level of functioning.

PRESENTING PROBLEM

Robert stated that he was in the hospital "for family problems." (What kind?) "Just not getting along. Not being able to work out at home." He said these problems started two years ago when "my mom put me into a private school, and I got kicked out" for "doing wrong things. Inappropriateness." Reportedly, on March 9, 2000, Robert consumed a large quantity of alcohol, became acutely intoxicated, and was agitated, belligerent, combative, tearful, and yelled obscenities. His family took him to Monte hospital where tests revealed a blood alcohol level of .24 without indication of other drugs. Seemingly, he consumed this near-lethal dose of alcohol in response to his hatred and fear of his parents.

BACKGROUND INFORMATION

The following information was obtained from Robert, who was a restless and dubious historian, from his hospital chart, and from a March 12, 2000, telephone conversation with Dr. Smith.

Robert was born and raised in San Francisco County and his biological parents remain married. The father, approximately age 40, is a career Air Force person with a rank unknown to Robert. He said his father teaches machine maintenance and attends college. He described his father as "a mean, violent, cruel person." (How does he show this?) "He used to hit me." (What would cause him to do that?) "Sometimes when I did something wrong." (Like what?) "Anything, any little thing, sometimes." He said he was hit with objects such as "a belt, sticks, spoons, forks, knives. Any object a lot of times." He said the most significant injuries he suffered during such beatings were "large blood blisters on my hand." He said he never suffered an injury that required medical attention but that he did ask his father's permission to remain out of school to conceal the marks from his peers. He said his father refused the request.

Robert's mother is approximately age 35 and a business secretary. Asked to describe her, he said, "I don't know. I guess she's mean. That's what I think. Is it my opinion?" (Yes. Mean in what way?) "Verbal abuse, yelling, screaming." (What causes her to do this?) "Anything, like my father. Frustration." (How does she react when your father hits you?) "She doesn't do anything. She lets him."

Robert is the oldest of three siblings with a brother, age 13, and a sister, age 10. Both siblings live in the home and attend school. "My brother's a little strange. Very silly and inappropriate for his age." He said his brother shows these characteristics when he "runs around naked, talks real strange like he's on an acid trip when he's not." He said his brother excels in school and has never received counseling or psychotherapy. Regarding the sister, "She's fine."

Medically, Robert believes he experienced an uneventful gestation and delivery. Regarding unusual illnesses or accidents, he said that at age 10 or 11 he broke either his wrist or his thumb during horseplay with friends and that the break healed without limitation. He said that at a later age he could not recall he broke his finger and wrist in horseplay and that he had broken such bones approximately four or five times. He said that otherwise his health has been good.

Asked about unusual events in his life before starting school, Robert said, "I can't remember that far back." He said he enrolled at Rutledge Elementary School at the usual age where he remained in regular classes through the sixth grade. He earned primarily "C's," and related to his peers "well." He said he also related to his teachers "well." He said he had no behavioral problems in elementary school and that his life was uneventful.

For the seventh grade, Robert was placed in a Roman Catholic school "because during the sixth grade they felt I was failing, not doing my work, taking too much leisure time." He said he remained at the Catholic school until the last month of the school year when he was expelled because of frequent misbehavior. Initially, he said he could not remember what act led to the expulsion but later said, "I lit a firecracker in class." (In your opinion, why did you do that?) "Rebelling." (Against?) "Parents. It was also for fun." (Why were you rebelling against them?) "Cause I didn't want to go to that school. I wanted to go to public school." (Why?) "Because I had gone to a public school for six years, and all my friends were going there."

Robert said he remained home without attending school until the following fall, September 1999, when he enrolled in the eighth grade at Franklin Junior High School to remain in regular classes until entering Monte Hospital. In the eighth grade, his marks were "failing, no, below standard." (Why were they so low?) "Because I had started using alcohol, drugs. I don't know. I don't know if that's the reason." (Why did you start using drugs?) "I don't know that, the reason why." He said he was introduced to alcohol and other drugs by his age peers. He estimated that he was intoxicated approximately 30 times during the eighth grade, usually at home with friends while his parents were out. The only street drug used was marijuana. He said his parents did not know about his substance misuse but complained about his grades. "They wanted me to bring them up, but I started failing even more." He said otherwise there were no unusual events during the eighth grade.

Last September, Robert enrolled in the ninth grade. Then, "I started hanging out with more friends that drank and used more drugs," but he misused only alcohol and marijuana. He said he was intoxicated "a few times a week" and misused marijuana "maybe four or five times a week." He said that otherwise his life was uneventful until he entered Monte Hospital for the first time during Christmas vacation because "I was dead drunk." He said he remained hospitalized three days, until "the doctor thought I

should be discharged." A week later, he returned to Monte Hospital for his current hospitalization.

Asked about his future plans in life, Robert said, "I haven't made any. I want to live with my grandparents, but they're not going to let me." He said his mother and his treating psychiatrist were against this plan because "they think I'm running away from my problems."

INTERPRETATIONS AND IMPRESSIONS

On the WISC-III, Robert obtained a Verbal IQ of 91, a Performance IQ of 84 and a Full Scale IQ of 86. His Verbal scaled scores ranged from 7 to 10 without significant deviation from the Verbal mean. Similarly, his Performance scaled scores ranged from 6 to 10. Verbally, he functioned at the 27th percentile and in the Average range. In terms of perceptual motor performance, he functioned at the 14th percentile and in the Low Average range. Overall, he functioned intellectually at the 18th percentile and in the Low Average range. His performance is probably a conservative reflection of his intellectual capacity, which may be as high as the high average range.

The structure of Robert's personality was difficult to estimate from his limited 14 Rorschach responses but appears nonpsychotic. Because of his limited responses, his ratios and percentages should be interpreted with caution. He uses repression and denial excessively in an attempt to exclude unacceptable impulses and fantasies from consciousness ($F = 92\%$). Nonetheless, Robert has difficulty seeing those things in the environment seen by most persons (Populars = 1) and at times gravely distorts his perceptions ($X + \% = 65\%$). These distortions are not frequent and are most likely to occur when he is emotionally stimulated. Because of his rigid defensive posture, he currently has few energies available ($M + C = 2.0$). He attempts to enhance control by denying emotional stimuli ($Sum\ C = 0$). And generally when expressing emotion, he has adequate control ($FC : CF + C = 1 : 0$). Because of his rigidity, his aspirations exceed his ability to perform ($W : M = 7 : 0$), and there is no indication of whether he prefers to accomplish more in fantasy or in action ($EB + 1 : 1$). Because of his excessive repression and denial, his depression is not obvious as he does not easily respond to emotional material ($Afr = .60$). For his age, his interests are somewhat narrow and immature ($A = 64\%$).

The content of Robert's personality projected in his Rorschach responses is a perception of himself as confused and disorganized, much as he projected in his first response to Card I: "A star; a messed up star. Someone drew it wrong." As he noted in his Rorschach scores, he attempts to deny emotional stimuli to enhance control, a mechanism typified in his rejection of Cards VII and IX, the first of three highly chromatic and, therefore, affectively stimulating cards. When presented with Card VIII, he responded in approximately three seconds, "I see nothing in this, I don't see anything." His response to Card IX was similar. On Card X he used the whole of the blot as an "insect," a response that had negative form level. Cautiously, he saw on each of the first five Rorschach cards the percept of "insect." Such perseveration is often found among neurologically damaged persons although it appears more likely in Robert's case a product of anxiety that limited his ability to structure the amorphous blot material much beyond an unspecified insect.

The content of Robert's personality as projected in his TAT responses is a painful perception of himself as evil, a perception that may have resulted in part from his already overly punitive superego receiving further stimulation while he was enrolled in a parochial school for the seventh grade. The content suggested by his cryptic response to Card 2 (man plowing with a horse in the background, a pregnant woman standing to one side near a tree, and a girl in the foreground holding books) was:

> Past, going, gone to a church school. Present, leaving home. Future, death. (Death in what way?) Starvation.

Robert views his internal anger and violence as undeniable indicators of his evilness, and evilness from which he can escape by drowning it along with himself in alcohol. Vivid were his responses to Cards 8BM (a primitive surgical scene with a boy standing in the foreground and a gun to one side) and 13MF (a bare-breasted woman lying on a bed and a man standing with one hand to his forehead):

> 8BM Past, Vietnam War. Present, remembering. Future, death from a heart attack. (What was he remembering?) The people being cut up in the Vietnam War.
>
> 13MF Past, alcoholic. Present, wife dead. Future, death from alcohol. (Why did the wife die?) Kidney disease.

Alternatively, he conceptualizes continuing to live with the evil in him without death as projected in his response to Card 15, a figure standing among tombstones:

> Past, possessed with evil. Present, trying to destroy the evil. Future, living with evil. (What was the evil?) The devil.

He believes that were he to continue living with his evil, he would represent a danger to other persons. This content was typified in his response to Card 7BM, a younger and older man in conversation:

> Past, jail. Present, courtroom. Future, jail. (Why was he in jail?) Murder.

Robert believes that in the past he was unwanted by a significant other. This awareness instilled in him the belief he was lacking in some way. When someone now offers him acceptance, he rejects it although he hopes he will be able to accept it at some point in the future.

On the Bender, all figures were completed with relative accuracy, and there were no indications of a neurological dysfunction affecting Robert's perceptual motor control. There were several indicators of impulsivity including overlapping, poor planning, and poor overall quality.

Asked to draw a picture of his family, Robert drew from left to right persons he later labeled verbally as "Dad," "Sister," and "Mom." Above these three and along the upper

margin of the page from left to right, he drew "Me" and "Brother." The drawing projects Robert's perception of him and his brother as separate from the parents and sister, much as he indicated during the interview when he said, "My brother's a little strange"; and clearly, he views himself as somewhat strange. Not surprisingly, he perceives the father as powerful and his mother as significantly weaker than his father.

DIAGNOSTIC IMPRESSIONS *DSM-IV*

AXIS I	300.40	Dysthymia, Secondary Type, Early Onset
	303.90	Alcohol Dependence, 305.20 Moderate Cannabis Abuse
AXIS II	v71.09	No Developmental or Personality Disorder Diagnosed
AXIS III		No Physical Condition or Disorder Diagnosed
AXIS IV		Psychosocial Stressors: Perception of himself as intrinsically evil and unlovable, believed rejection by parents, effects of drug misuse, failure to adjust socially and psychiatric hospitalization. Severity: 4-Severe (Admixture of acute events and enduring circumstances)
AXIS V		Global Assessment of Functioning (GAF): Current GAF: 35 Highest GAF past year: 50

SUMMARY AND RECOMMENDATIONS

The previous information suggests the following answers to the referral questions:

- Robert functions intellectually in the low average range although his endowment may be in the high average range and is currently reduced by his mental condition. He does not suffer from a neurological dysfunction affecting his perceptual motor control.
- His personality structure is nonpsychotic and his difficulties are more consistent with dystyhmia with alcohol dependence and moderate cannabis use.
- His personality structure is exceedingly rigid and characterized by repression and denial such that he fails to see those things in the environment seen by most persons and has few energies available. He attempts to enhance control by denying emotional stimuli and generally can maintain control when responding emotionally. When highly stimulated emotionally, however, he gravely distorts his perceptions. With his rigid defense structure, he has few energies available to gratify his aspirations but is capable of concealing overt indications of depression such as tearfulness and significant withdrawal. When he is consuming disinhibiting chemicals such as alcohol, his behavioral and emotional expressions are far less controlled.
- Robert has at least a vague awareness he is "messed up." It is probable he was highly critical of himself before enrollment in the Roman Catholic school. There he may have been overly scrupulous when interpreting moral teachings, and his already punitive superego became self-destructive. He may have felt most guilty

about the commandment to honor parental figures. He came to view himself as anathema. He believes his alternatives are to continue living with the believed devil inside him with the probability he will eventually murder someone or, alternatively, to end his life to escape the internal evil and avoid harming others. Not unexpectedly, death is frightening to him and presently he flirts with self-destruction by poisoning himself with alcohol.

A more complete social history would be useful in determining the major factors contributing to Robert's strong negative self-image. Although it is unlikely Robert's father is as abusive as Robert describes him, some series of events have led Robert to view his father as unaccepting and unacceptable. A combination of family psychotherapy on a weekly basis and continuing individual psychotherapy with Robert are the treatments of choice. Initially, Robert will find individual treatment threatening and, consequently, may gain more in the family sessions. As he progresses in treatment, he will profit more from the individual sessions. Without treatment, he is a significant danger to himself and could end his life in suicide.

THE LEGAL CONTEXT

Dear Mr. Hamlin:

Thank you for the opportunity of commenting on the March 21, 2000 report of James A. Speculus, MD, undertaken as an evaluation of Mrs. Sharon Paine. As you know, Mrs. Paine is a 35-year-old, Caucasian married female with a university education who sustained a motor vehicle accident on May 17, 1998. She experienced injuries to her mouth and left shoulder that have caused her ongoing physical pain, several surgeries, and emotional distress. In addition, she has evidently had difficulties adjusting to her condition, which has precipitated symptoms of insomnia, depression, feelings of hopelessness, irritability, anxiety, and physical pain. According to the March 21, 2000 report of Dr. Speculus, these symptoms had mostly been resolved at the time of his assessment and he concluded that the complaints were mostly due to malingering. It is my understanding that you would like me to comment on the methodology in Dr. Speculus' report with specific questions related to the validity of the Test of Memory Malingering (TOMM), the appropriateness of using the TOMM with the client, an optimal protocol for assessing malingering for a client such as Mrs. Paine, possible reasons why Mrs. Paine would have scored poorly on the TOMM, and the training/qualifications required for a person using a test such as the TOMM. I note in this regard that I was neither requested to actually evaluate the client's condition, nor did I actually meet with her directly. Rather, I will contain my comments to the methodology and instruments used by Dr. Speculus.

For clarity and ease of reference, my comments are numbered:

1. The TOMM is essentially a well-designed, psychometrically sound test designed for assessing the exaggeration of memory deficits among persons with known or suspected neuropsychological impairment. By this, I mean that it is fairly accurate in

distinguishing persons simulating memory deficits from those who are normal and even those who have well-documented, physically based memory deficits (Rees, Tombaugh, Gansler, & Moczynski, 1998; Tombaugh, 1997). Thus, it is a good test for the purposes for which it was designed. However, one of the weaknesses of the TOMM, as stated by researchers, is a "failure to include a psychiatric control group to evaluate the possible effects of psychiatric symptomology, such as depression, on the TOMM" (Rees et al., 1998, p. 18).

2. In evaluating the appropriateness of the TOMM for Mrs. Paine, it is essential to point out that Mrs. Paine was presenting with and being evaluated for symptoms of pain, depression, and anxiety. Thus, it is confusing why a test designed to assess the possibility of exaggerating *memory* deficits was used. Indeed, I note that there is an extensive literature and a number of well-developed tests that have been specifically designed to detect the possible exaggeration of emotional (depression and anxiety) difficulties. Some of these tests have the advantages of not only assessing for the validity of symptoms, but can also assess a wide range of highly relevant emotional and personality variables. Thus, I would have thought that Dr. Speculus would have tailored his test selection based on the presenting problem of the client. I conclude then, that although the TOMM is indeed a good test for the purposes for which it was designed, in this case it was not being used for the purpose for which it was designed.

3. A further issue is that a single test, particularly one such as the TOMM, should never be used in and of itself to draw final conclusions about a client. Standard texts on assessment all stress the importance of integrating multiple sources of information in making final conclusions (see Beutler & Berren, 1995; Groth-Marnat, 1999, 2000a). In contrast, the format of Dr. Speculus' March 21, 2000 report suggests that he used a mechanical, unintegrated, test-dominated manner of making his conclusions ("I asked her to perform the TOMM. She scored 35 out of 50 . . . this indicates an intention to deceive me." p. 6). This issue of integrating converging sources of information is particularly crucial in the case of malingering where a conclusion that the client has malingered can have a major impact on the client's life. A protocol for assessing malingering as recommended by major publications in the field (see Blau, 1998; Franzen & Iverson, 1998; R. Rogers, 1997) would include not only information on a single (appropriately used) test and the context of possible gain, but also relevant dimensions such as personality (i.e., antisocial trends), attitude (i.e., "the world owes me a living"), past history suggestive of using people/organizations for gain, collateral interviews with outside sources, multiple observations of the client's behavior, and patterns among several test results (i.e., inconsistent patterns between tests). My reading of the report indicates that few of these dimensions were evaluated. (Although there was a comment that "she was careful about choosing her words," care in wording would be expected for the majority of people being evaluated, particularly a university-educated person such as Mrs. Paine.).

4. Despite the inappropriateness of using the TOMM and its apparent lack of integration with additional relevant (mostly unassessed) aspects of the client, it still remains unclear as to why she scored as low as she did. Possibilities include an undiagnosed neuropsychological condition, exaggeration of resolved (but currently minimal) difficulties, exaggeration of actual quite significant difficulties, or the impact of emotional factors on test performance. Given her age and history, an undiagnosed neuropsychological condition such as dementia is unlikely. A further option (taken by

Dr. Speculus) is that she was exaggerating minimal (but now "resolved"; see p. 6 of report) difficulties. In contrast, another option might be that there was some exaggeration, but there were also quite significant difficulties the client was experiencing. Indeed, the context and process of medico-legal evaluation sometimes encourages clients with actual difficulties to make sure their difficulties are noticed by the client exaggerating these difficulties. This can clearly complicate the process of evaluation and make assessment of the true extent of difficulties quite challenging. I note in this regard that Dr. Speculus' report did not seem to consider any of these options. I would speculate that, given the recent death of Mrs. Paine's father (4 to 5 weeks before the March 21, 2000 report), shoulder injuries, and ongoing palate pathology, it would be unusual if Mrs. Paine were not experiencing a relatively high amount of distress. Finally, as indicated in the Rees et al. (1998) study, cognitive tests can be lowered by emotional factors (e.g., Burt et al., 1995; Finlayson & Bird, 1991; Sbordone, 2000a). These last two possibilities (exaggeration but with actual significant underlying difficulties and lowering of test performance due to emotional factors) seem to have been insufficiently explored as possible reasons for the lowered TOMM performance.

5. Training in the appropriate use of psychometric instruments involves a relevant professional degree, minimum of two full semester units in psychological testing, a semester on statistics, additional relevant course work (personality theory, psychopathology, psychotherapy, etc.), and two years of supervised experiences of which a substantial portion should relate to the use of tests with actual clients. Key features of this training would emphasize the importance of evaluating a test's norms in relation to whom the test will be used on, tailoring tests for the specific presenting problem, developing an optimal assessment protocol based on relevant research, and integrating test scores with a wide range of relevant information (see Groth-Marnat, 1999). I note that test developers do screen potential purchasers of restricted tests (such as the TOMM) by using a brief series of questions. However, these brief screening questions do not guarantee that this minimal training has actually been undertaken.

In summary, the assessment of Mrs. Paine did seem to elicit a number of seemingly relevant aspects of her history. However, the assessment of possible malingering was not optimal because:

- The test selected (TOMM) was not tailored toward her presenting problem.
- The test was not normed for clients similar to Mrs. Paine.
- The interpretation of the TOMM appeared to be done in a mechanical, test-dominated manner.
- Alternative possibilities for the meaning of Mrs. Paine's score on the TOMM did not appear to be sufficiently explored.
- The assessment of malingering appeared to include only a minimum number of relevant variables.

I hope that the previous perspective and comments help to better evaluate Mrs. Paine's case. If you have any questions or would like further elaboration, please contact me at your convenience.

THE EDUCATIONAL SETTING*

NAME: Anna S.
DOB: 05/21/89
GENDER: Female
DATES OF ASSESSMENT: July 17, 21, 22, 24, 2001
DATE OF REPORT: August 4, 2001
EXAMINER: Annie Chung, Ph.D.

REFERRAL QUESTION

Anna is a 12-year, 1-month-old bilingual, Hispanic female who is currently residing in a foster home due to parental abuse, neglect, and abandonment. Anna presents with irritability, anxiety, and poor peer relations. Additionally, Anna's foster parents indicate that Anna is having problems in school because of interpersonal difficulties with her classmates and problems adjusting to their home (e.g., seeks constant attention, argues with foster siblings). Anna's therapist, Mary Smith, LCSW, requested this psychological evaluation to assist in assessing her current level of cognitive and emotional functioning, clarify her diagnosis, identify her strengths, and provide suggestions that may be useful in helping her to adjust to school.

EVALUATION PROCEDURES/SOURCES OF INFORMATION

Clinical interview with Anna.

Review of social service/family court documents.

Review of current case file.

Consultation with Anna's social worker, therapist, teacher, and foster parents.

Wechsler Intelligence Scale for Children-III (WISC-III).

Wechsler Individual Achievement Test—Screener (WIAT).

Trauma Symptom Checklist for Children (TSCC).

Personality Inventory for Youth (PIY).

Children's Apperceptive Test.

Draw-A-Person/House-Tree-Person.

Child Self-Report and Projective Inventory (Color How You Feel, Critical Items, Color How Others Make You Feel, Perceived Competence, Draw a Child in the Rain, Sentence Completion, Kinetic Family Drawing).

BEHAVIORAL OBSERVATIONS/MENTAL HEALTH STATUS

During the initial testing session, Anna presented as a pleasant and engaging youth who was casually dressed and appropriately groomed. She appeared to be of average size and weight for her gender and chronological age. She willingly accompanied this examiner and appeared eager for individualized attention. Anna remained cooperative, maintained good eye contact, and readily responded to questions with descriptive

* Report submitted by Annie Chung, PhD.

details. She was clearly oriented to person, place, time, and situation. Her mood was somewhat depressed and her affect was generally constricted. She indicated that the reason for this evaluation "is to see how I'm doing." Anna demonstrated fair insight regarding her current situation, identifying her mother's abandonment of her as the reason for her current foster care placement and her associated feelings of distress. Her responses to hypothetical situations requiring decision-making skills suggest somewhat compromised judgment. Thought content was appropriate to the situation and thought processes were lucid, concrete, and coherent. There was no evidence of perceptual disturbances, flight of ideas, circumstantiality, or loose associations. Cognitive abilities appeared to be within the average range. Although not formally assessed, her conversational English skills were apparently proficient. When responding to direct questions, Anna typically responded in grammatically correct English. At times, she responded in Spanish but quickly restated her response in English, switching between both languages with apparent ease. During subsequent testing sessions, Anna's test-taking behaviors continued to be cooperative. Her approach to each task was marked by diligence, good effort, eagerness to please the examiner, and heightened performance anxiety (e.g., "Am I doing a good job?"). Although Anna exhibited good attention and concentration with the majority of testing activities, she also worked rapidly when given written tasks. Anna reported that she is "good at writing fast" and stated that she wanted this examiner to see how quickly she can work.

Even though Anna appeared to give her best efforts to the tasks presented to her, she is a bilingual youth whose first language is Spanish. Given that the tests administered were not specifically normed on this population, the assessment results should be treated with caution. In addition, she may have somewhat overreported her current level of emotional difficulties.

BACKGROUND INFORMATION

Information obtained from family court documents, consultation with Anna's therapist, her social worker, her current foster parents, and interview with Anna indicates that Anna has experienced multiple disruptions in her home life events throughout her childhood. These events include physical/emotional abuse, neglect, and abandonment. Anna lived with her biological mother, Lupe S., until she was approximately 6 years old. At that time, Anna and Ms. S. moved in with Juan F., whom Anna refers to as her stepfather. Anna continued to reside primarily with her mother, stepfather, and their two children, Roberto and Juanita, until March 2001.

After moving in with her stepfather, Anna reported that her mother repeatedly stated, "I wish you [Anna] were never born," and that her parents frequently yelled at her, hit her, and did not provide adequate clothing or care for her. Furthermore, Anna reported that she and her mother frequently left her stepfather's home to live with other relatives or in homeless shelters because of her stepfather's "mean" behaviors but that her mother always returned to him. Family court documents indicate that there were five substantiated reports of physical/emotional abuse and neglect of Anna and/or her siblings by both her mother and stepfather. Anna's parents reportedly did not comply with family supervision plans as mandated by Child Protective Services. Additionally, Anna explained that over the past several years her mother exhibited "weird" behaviors such as rummaging through garbage cans, forgetting to wash clothes, and most pertinent

to the child, witnessing her mother's attempt to choke herself. Anna also reported that her mother has been hospitalized at least 4 to 5 times for attempting to kill herself. In the past year, Anna was temporarily placed at the local children's shelter after running away from her home and when her stepfather attempted to abandon her at this facility. Subsequently, Anna resided with a maternal aunt for an undetermined amount of time before returning to her mother's care in the spring of this year.

During the spring of 2001, Anna and Ms. S. resided intermittently with relatives, Ms. S.'s employer, and then in a homeless shelter. Ms. S. then requested that acquaintances from church care for Anna. The acquaintances agreed to this on a temporary basis. However, they reported that Anna began having significant problems at school (e.g., suspended for fighting). Furthermore, these caretakers stated that Anna appeared to have significant "emotional problems" and were, therefore, unwilling to care for her any longer. They contacted Social Services and explained Anna's situation, clarifying that her mother was not available to care for her. As a result, Anna has been a ward of the state since May 2001, and has been placed in the foster home of Mr. and Mrs. G.

Anna's social worker referred Anna for mental health services given her history of neglect and abuse, physical aggression in school, and her excessive irritability. On placement in her current foster home, Anna initially appeared very sad and had difficulties interacting with her foster siblings. Anna's foster parents indicated that Anna has not exhibited overt behavioral problems but does appear to be immature (e.g., excessive teasing, difficulty sharing), needs a considerable amount of individual attention, and displays excessive moodiness.

INTERPRETATIONS AND IMPRESSIONS

Cognitive Functioning. On the WISC-III, Anna achieved a Full Scale IQ of 85; Verbal IQ of 87; and Performance IQ of 86. Overall, this places her in the Low Average range or 16th percentile (lower 16% of the population) when compared to her age-related peers.

A close examination of Anna's relative strengths indicates relatively intact visual-spatial/mechanical skills (Perceptual Organization Index = 91) as well as good verbal expression/conceptualization skills associated with school-related learning (Verbal Comprehension Index = 89). She is likely to do particularly well in tasks requiring her to understand social norms and follow the meaning of nonverbal social situations. This was consistent with this examiner's observations of Anna during the interview in which she responded appropriately and answered all questions clearly.

In contrast to these strengths, Anna's responses to tasks that required her to respond quickly and accurately to nonverbal information were quite low (Processing Speed = 67). This suggests that even though she can understand what might be expected of her in school, she will do poorly when she is required to perform under time constraints. In these conditions, she would be likely to sacrifice speed for accuracy. In addition, she may have a difficult time concentrating, especially when dealing with numerical information presented to her verbally (e.g., recalling phone numbers, making arithmetical calculations).

In comparison to her low average global cognitive potential (WISC-III FSIQ = 85), Anna's composite academic achievement skills (WIAT Screener; Total *SS* = 93) is

within the average range. Anna's scores on the Spelling (*SS* = 102) and Basic Reading (*SS* = 98) subtests suggest that these are relative strengths while her Mathematical Reasoning (*SS* = 81) skills are considerably weaker. However, Anna's solid average performance in Spelling and Basic Reading, which actually exceed her global cognitive potential, suggests that Anna's scores on the WISC-III may be an underestimate of her true cognitive potential. This is likely because of her bilingualism and the absence of normative data pertaining to this specific population's performance on standardized intellectual assessment measures. Her lower scores on the IQ tests are also likely to result from her difficulties working under time constraints.

Emotional Functioning. Anna's reports of multiple and atypically severe intrapersonal discomfort is associated with her traumatic and disruptive childhood experiences. She is experiencing heightened levels of internal anxiety, depression, and uncertainty as well as low self-esteem. For example, she frequently provided self-deprecating remarks about her drawings (e.g., "it's ugly"), and the colors she used to represent her feelings depicted her emotional states as overwhelmingly sad and worried. She seems to harbor feelings of anger but has difficulty expressing these appropriately and, thus, may have a propensity to act out. However, it also appears that she attempts to mask her internal distress by presenting herself as outwardly happy and content. Furthermore, Anna depicts her life experiences as extremely stressful and has significant difficulty coping with such experiences. Her perceptions of her family and others indicate significant ambivalence, isolation, and alienation. For example, when prompted to draw a picture of her family, Anna did not include herself and depicted all family members enjoying activities independent of each other. However, she also incorporated symbols of warmth and closeness in her house drawings. Thus, Anna appears to hold conflictual feelings about her family while she also hopes for safety and support.

In response to questions eliciting Anna's experience of critical life events, she identified three specific traumatic experiences that appear to be associated with her current worries, fears, nightmares, and sadness. Anna reported that when she was about 8 years old, a homeless man attempted to take her while she was walking by herself in the evening but that she never shared this experience with her parents because she felt that it would not be important to them. Additionally, Anna explained that when she was about 10 years old, she witnessed her mother attempting to choke herself and subsequently worries that her mother will kill herself. Most recently, Anna reported that her mother left her with people that she did not know and has been feeling very sad and worried since that time. Anna indicated that she has difficulty staying asleep throughout the night as she has bad dreams about these "scary" events. She also admitted to having considerable difficulty concentrating at school because of intrusive thoughts about her past experiences and worries about her future.

Anna's interpretation of story cards was representative of overwhelmingly negative outcomes for the child figure. Themes included betrayal by mother, excessive worries and fears, abandonment and rejection by others, lack of caring and attention from mother, foreshortened future, harm inflicted by others, and inability to obtain wishes and desires such as reunification with family. Thus, it appears that Anna experiences her world as generally unsafe, unpredictable, and lacking emotional warmth. However,

in one story, Anna projected a positive resolution for the child, in which the parents comforted and cared for the child "when she needed them." This suggests that Anna has a strong desire to be nurtured and accepted by caregivers.

In spite of these traumatic and stressful life experiences, Anna is able to identify positive self-attributes related to her academic skills (e.g., spelling, handwriting), personal appearance, and athleticism. It appears that these strengths and perceived competencies bolster Anna's self-esteem and promote her ability to function on a day-to-day basis and hope for positive changes in her life.

Interpersonal Functioning. Anna typically responds to people with the hopes of obtaining attention and acceptance. This has been evident in her interactions with adults, which have been characterized by a strong eagerness to please, requests for individual attention, and gift-giving. However, Anna's negative experiences with her parents have likely bolstered her defenses against continued rejection. Therefore, Anna may respond by rejecting others first. This has been exemplified in her history of aggressive peer interactions, which have led to persistent teasing and fighting at school as well as her own reports of not having any "true" friends. While such behaviors have not been problematic at her current placement, Anna's foster parents have reported that her social interactions with peers have been immature (e.g., whining, teasing, difficulty sharing). Yet, when provided with support and nurturing, Anna has become increasingly able to verbalize her difficulties and share painful experiences with adults (e.g., therapist, foster mother).

Behavioral Functioning. Assessment of Anna's overt behavior did not reveal high levels of acting out (e.g., aggression, delinquency). In contrast, she is likely to internalize her distress by becoming anxious and depressed. Anna's foster mother, Mrs. G., described Anna as initially very argumentative and oppositional but said that these behaviors subsided within the first several weeks in the new home. Overall, Mrs. G.'s ratings of Anna's current functioning suggest that Anna is not displaying significant behavioral concerns but that she does demonstrate symptoms of emotional distress at this time. It appears that Anna is responding positively to the structure, consistency, and nurturing provided by her foster parents although concerns associated with Anna's internal turmoil are evident.

DIAGNOSTIC IMPRESSIONS

AXIS I	309.81	Posttraumatic Stress Disorder
	995.5	Neglect/Physical Abuse of Child
AXIS II	v71.09	No Diagnosis on AXIS II
AXIS III		None known/reported
AXIS IV		Abandonment by mother
AXIS V		Current GAF: 60 Highest GAF past year: 50

SUMMARY AND RECOMMENDATIONS

Findings from this psychological evaluation indicate that: (a) Anna's academic skills in reading and spelling exceed her measured cognitive abilities, (b) she does not currently exhibit overt behavioral difficulties although peer/adult interactions have been strained, and (c) she is demonstrating and experiencing significant signs of emotional distress. While Anna's global cognitive ability score is within the low average range, it is likely an underestimate of her true ability level given her higher than expected scores on spelling and reading skills. This suggests that Anna's bilingualism and slowed performance may have inhibited her ability to demonstrate her optimal cognitive ability on the WISC-III. It is also quite possible that Anna's exposure to early and persistent abuse/neglect may have impeded the development of her optimal cognitive functioning, which is likely to be in the average range. Anna's experience of multiple traumatic events, including the threat of an abduction by a stranger, witnessing her mother's suicide attempts, and abandonment, have further exacerbated her feelings of fear/uncertainty associated with persistent maltreatment by her parents. It is evident that Anna's worries, sadness, difficulty concentrating, and recurring nightmares are associated with these traumatic experiences. Subsequently, Anna's internal turmoil has been increasingly evident in her interpersonal relationships. With adults, Anna exhibits a strong desire to please and seeks constant individual attention. With peers, her attempts to be recognized have included a large repertoire of negative attention-seeking behaviors. Although obstructive, these patterns of relating to others appear to represent Anna's attempts to obtain any type of personal acknowledgment, albeit negative, in a world that she perceives as largely unsafe and threatening.

However, it is also important to highlight Anna's resiliency in the face of immense risk factors in her young life. Anna's ability to acquire new knowledge (e.g., proficiency in English) and her attempts to initiate relationships with others suggest that she possesses the capability to succeed in academic settings and is receptive to building social support networks outside of her family of origin. In spite of her current and expected difficulties associated with trauma, Anna presents as a likeable youth who desires acceptance from others. Ensuring placement in a loving, consistent, and structured home environment that provides unconditional acceptance is absolutely crucial in supporting Anna's cognitive and socioemotional development and long-term stability. This is also likely to improve her cognitive level of functioning and enhance her academic performance.

Given the previous findings, the following recommendations seem appropriate:

1. Ensure regular attendance at school to bolster Anna's academic skills. Providing extra assistance/tutoring for mathematics and strategies on how best to work with timed tasks would be particularly helpful.
2. Provide Anna with opportunities to participate in structured community activities of interest to her (e.g., sports team) to promote her sense of self-efficacy and enable her to develop positive peer relationships in natural settings.
3. Provide Anna's foster parents with ongoing parenting education and parenting skills training to promote their understanding of Anna's needs (e.g., supportive structure, limit setting, explanations, consistency, acceptance) and use of

specific strategies to promote Anna's positive adjustment to her new placement, as well as school.

4. Reevaluate Anna's cognitive and academic functioning, including language proficiency in English and Spanish, after stable home and academic placements are secured, to determine if a need for additional educational interventions are warranted.
5. Provide Anna with individual therapy that uses supportive as well as cognitive-behavioral strategies to enhance her ability to further process her thoughts/feelings about her situation, promote her self-esteem, facilitate her use of positive coping strategies when distressed, and alleviate her trauma-related symptoms. As therapy progresses, it would be important to help Anna understand and process her experience of her family, her abandonment, and abuse. Integrating and resolving these issues will be essential in assisting her to become a well-adapted adolescent and young adult.

THE PSYCHOLOGY CLINIC*

NAME: A. G.
DATE OF BIRTH: May 30, 1925
DATE OF EXAMINATION: December 12, 1995
CASE #: 96–041
SEX: Female
ETHNICITY: European American
REFERRED BY: Dr. M.

REFERRAL QUESTION

This 70-year-old, divorced woman was referred for psychological evaluation by Dr. M., who specified that the patient suffered from agoraphobia and requested assistance in identifying effective treatment for her condition. A. G., on the other hand, indicated that she had had agoraphobia in the past, but described her current problem as one of motivation rather than panic or fear. She asserted that she is currently not immobilized nor is she extremely anxious when she travels. She attributes her restrictive lifestyle to the absence of "energy" and "motivation" to travel and engage in social activities. She acknowledged an "underlying apprehension" that arises when she is scheduled to leave home, however, resulting in her putting off her departure as long as possible. After she actually goes out, she reported that she remains anxious until she returns home.

The current evaluation was designed to clarify the nature of the problem, to develop treatment plans, and, if indicated, to initiate a treatment program.

EVALUATION PROCEDURES

By prior arrangement, A. G. completed the Life History Questionnaire before her intake interview on December 12, 1995. She was unable to complete additional testing

* Report contributed by Larry Beutler, PhD, Pacific Graduate School of Psychology, Palo Alto, CA.

during the intake because of "discomfort" and "apprehension" about making the drive home. She was, therefore, rescheduled with the request that she complete the rest of the paper-and-pencil materials over the course of the next two weeks.

During the total assessment sessions, she was evaluated using the following procedures:

Life History Questionnaire (12/12/95).
Clinical Interview (12/12/95).
Structured Clinical Interview for *DSM-IV* (SCID; 12/22/95).
State-Trait Anxiety Inventory (STAI; 12/13/95).
Millon Clinical Multiaxial Inventory-III (MCMI-III; 12/13/95).
Personal Attitude Inventory (Dowd Therapeutic Reactance Scale; 12/13/95).
Beck Depression Inventory (BDI; 12/13/95).
Beck Hopelessness Scale (BHS; 12/13/95).
Minnesota Multiphasic Personality Inventory-2 (MMPI-2; 12/21/95).
Sarason Social Support Questionnaire (SSQ; 1/2/95).

BEHAVIORAL OBSERVATIONS

Throughout the evaluation, A. G. was articulate, introspective, and cooperative. While acknowledging discomfort in talking about her sexual orientation and religious feelings, she was quite forthcoming when questioned directly. A valid MMPI-2 profile, along with her willingness to cooperate and introspect, suggests that the current evaluation presents a valid picture of her current level of functioning.

RELEVANT HISTORY

History of Presenting Problem. This 70-year-old, Jewish woman identifies three major symptom clusters that have been problematic for her: agoraphobia, panic attacks, and dissociation. She has a long history of panic attacks without agoraphobia, dating to age 12. The first panic episode occurred when she was babysitting for a family friend. She suddenly hyperventilated, began experiencing heart palpitations, and became afraid that she was going to die. She ran into the street yelling for help, but no one heard her or tried to assist her. The situation was resolved by exerting "self-control."

After her initial panic attacks began, they gradually increased to a frequency of about once per week throughout her teenage years. To protect herself from feared panic and what she perceived as possible death, she frequently slept with her parents and confined herself to known places and locations. At their worst, her panic attacks involved physical symptoms such as nausea, shortness of breath (hyperventilation), and dizziness, as well as cognitive symptoms such as fears of losing her mind, dying, of being overwhelmed, and unspecified danger. However, she learned to control these symptoms over time by avoiding such activities as going out, driving, and socializing with groups. These efforts have been successful in that A. G. reported that she had been asymptomatic for agoraphobia and panic for 31 years.

She currently reports that she has become apprehensive about travel and social activity, but the symptoms are confined to initial anticipatory anxiety, gastrointestinal distress, and headaches but with no heart palpitations, shortness of breath, or fainting. She prevents more extensive symptoms by avoiding travel and through a variety of distraction procedures. When she begins to experience the onset of panic, she calls someone or begins to read an interesting book. Her contacts with other people at these times do not include a disclosure of or discussions about the panic, but are reported to be simply methods to involve herself with others and to take her mind off her feelings.

She reported that the current symptoms are mild in intensity and include a general distaste of travel, an inability to get comfortable with being out alone, and a general heightened sense of vulnerability and apprehension until she is able to return home. She continues to avoid night travel and avoids being alone, if possible, to prevent the associated anxiety.

Since 1986, several dissociative episodes have occurred, which she believes were precipitated by her decision to openly acknowledge her homosexual orientation. The first instance followed a sexual encounter with her current partner during a vacation. After the sexual act, the client experienced an apparent fugue state. She became disoriented, was unable to recall personal information such as that her parents were deceased, and engaged in distraught communication with her lover about "Why am I here." The episodes have subsequently recurred several times: They come on suddenly and without warning and she subsequently has no memory for the events. They uniformly follow a lesbian sexual encounter, and if her partner remains with her during this period (sometimes up to several hours), the fears gradually subside. However, after these dissociative states, she reported having a sense of helplessness, hopelessness, confusion, headaches, and nausea that sometimes lasted for several days. She has been able to successfully avoid these episodes by not engaging in sexual activities for nearly six years.

History of Treatment. Ms. G. was first treated and hospitalized in 1965 because of agoraphobia. There have been no subsequent hospitalizations. However, she has entered into two treatment relationships in the years since. Her current medication is managed by a psychiatrist, who is treating her with Xanex, Tagamet, and Paxil. She reported that since being on the medication, she has been excessively sleepy and has a difficult time staying awake during the day. She also has experienced an increase in stomach difficulties and diarrhea. She was also treated for a short time in 1985 by an internist and psychiatrist. At that time, she was given tricyclic antidepressants. These drugs produced hallucinations and were discontinued shortly after initiation.

A review of this woman's symptom history also reveals that she has had substantial periods of time in which she has been asymptomatic for fugue states, panic attacks, and agoraphobia. She reported that between August 1994 and June 1995, she was the "best ever." She was able to travel alone, found life enjoyable, and experienced no episodes of discomfort or fear. More recently, she has gradually become more depressed and dysphoric as well as fearful, although there was no obvious precipitator for these feelings.

Family Background. A. G. was raised in a middle-class, Jewish family. She was the older of two children, having a brother who is one and a half years her junior. The

family always maintained at least a superficial religious identity and a facade of happiness. However, she reported that, behind this façade, there were significant underlying family conflicts. Religion has always been a source of conflict for her as has her sexual orientation. Moreover, she reported a long family history of mental illness and interpersonal conflict. While she described her parents as emotionally stable, both her parents' families have histories of psychiatric disorders. Her mother was the oldest of nine children; an uncle died in a halfway house with a diagnosis of schizophrenia, another was diagnosed as having bipolar disorder, and still another was diagnosed as having depression. On her father's side, at least one uncle is reported to have had a major depression that was treated with antidepressants.

In her own personal history, A. G. reported that she always felt confused about her sexual orientation. At age 17, she received a proposal of marriage. She declined but he persisted, and she went to live with an uncle to escape his advances. He pursued her and finally, against her "better judgment," he talked her into marriage. The newlyweds moved to Metroville to be with her family, but problems persisted and they separated after about a year. By that time, she had given birth to a daughter. She moved in with her parents, but long-standing conflicts with her mother became more frequent. When her husband contested and prevented the culmination of the divorce, A. G. moved out of the family home and went to work, leaving her daughter with her mother. A. G. blamed her parents for her failed marriage and refused further contact. She did not see or speak with her daughter for two years.

After a period of estrangement from parents and daughter, the patient was contacted by her attorney, who informed her that A. G.'s mother could no longer raise the baby. Her husband was also informed, and he demanded that A. G. reconcile with him to raise the child. A. G. agreed to do so if they would move to Betterville to make a fresh start. Shortly after moving, A. G.'s husband became disillusioned and returned to Metroville, leaving her to raise the child. It was very shortly thereafter that she acknowledged to herself that she was a lesbian. She subsequently engaged in a series of brief lesbian affairs, and adopted a "secret life" in which she prevented her parents and husband from an awareness of this emerging sexual orientation.

Still being unable to raise her daughter and work, the client gave up the child to a foster family for temporary care. After a few years, she initiated an effort to again assume care of the child. Concerned about raising the child in a lesbian relationship, she accepted the proposal of marriage from a man who knew about her lesbian lifestyle. He, nonetheless, agreed to adopt the child and allowed her to continue her lesbian affairs. Their marriage lasted 23 years and produced two sons. Although unsatisfied with her dual life, she waited until her younger son graduated from high school before she left the marriage and began to pursue lesbian relationships exclusively and openly. She met her current lover in 1986. This relationship continues to be close although they ceased sexual contact approximately six years ago to prevent dissociative episodes.

A. G. reported being close to her brother when she was young, but that relationship became disrupted during their teenage years. At that time, he had more problems with their parents than she did; and, because of these conflicts, A. G.'s parents left all their money to her rather than to her brother. However, she had come to feel alienated from him and subsequently refused to share her inheritance. In spite of having become quite wealthy, he became resentful of her refusal to share the inheritance and cut off the

relationship with her. In the past few years, she has tried unsuccessfully to reestablish this relationship. He has indicated his willingness to "forgive" her, but denies any desire for a continuing relationship.

Medical History. A. G.'s medical history is unremarkable. She currently takes Tagamet for stomach distress and Xanex for anxiety. Aside from some loss of hearing and psychophysiological symptoms, she acknowledges no significant medical problems.

INTERPRETATIONS AND IMPRESSIONS

While a formal assessment of intellectual level was not undertaken, both A. G.'s verbal conceptual skills and oral presentation suggest at least average and probably bright normal intellectual performance. Her ideation is dominated by preoccupation with ways to avoid uncomfortable feelings, along with concerns with physical symptoms. Collectively, this results in mild impairment to her cognitive efficiency. Her verbal processes are organized, circumstantial, and occasionally dominated by topics about which she has pressing concerns; but they reflect no disorganization, no memory impairment, and moderately well-developed associative and abstract reasoning processes. While she is oriented in all three spheres and manifests no significant mental impairment, she notes having always been concerned with the potential loss of mental functions.

A. G. denies dysphoria, depression, and anxiety. She complains of poor sleep, loss of energy, and lack of motivation. Formal assessment confirms the presence of vegetative signs (increased appetite, variable sleep, social withdrawal, loss of interest, reduced libido), consistent with the presence of mild to moderate depression without subjective dysphoria. Trait anxiety levels are within the normal range for her age, and subjective depression is only mild, with the dominant symptoms being psychophysiological. Her affect is appropriate, though somewhat variable. Affective responsivity is both dysthymic and blunted.

Ms. G.'s mood disturbance reflects a chronic condition, against which she has constructed a variety of rigid and brittle defenses. She is excessively sensitive to environmental signals of threat and, at the least suggestion of emotional arousal, engages in both direct and cognitive avoidance patterns. The result is that she prevents the intensification or even emergence of feelings that might overwhelm her. While protecting her somewhat from subjective sensations of anxiety and dysphoria, A. G.'s defenses are not sufficiently strong to prevent the emergence of a variety of secondary symptoms. Denial, phobic avoidance in the face of anticipatory cues, self-criticism, compartmentalization, and somatization are among her most frequently used defenses. As threat intensifies, her fragile denial deteriorates, and both somatization and direct avoidance predominate. Thus, acute stress evokes a variety of stress-related somatic symptoms and phobic behaviors that provide expression for her denial of anxiety and depression.

A. G. experiences ambivalent personality organization, with moderate disturbances to her functional adaptation. Her dominant conflicts involve strong needs for dependency, counterbalanced by equally strong strivings for self-definition. Her coping style involves both passive and active efforts to reconcile these strong drives. Thus, while she seeks approval and confirmation from others, even to the point of excessive subservience in which she gives up personal strivings, this is frequently a futile effort to ensure the

presence of other people in her life. Indeed, these efforts are usually designed to compensate for a host of covert rebellious and angry impulses and by overt efforts to be autonomous and self-guided. Thus, efforts to achieve self-fulfillment and autonomy are followed by guilt, self-doubt, and shame in which fear and withdrawal dominate. These latter symptoms, however, may be so demanding of attention that they are the functional equivalent of interpersonal anger, hurt, and resentment. Thus, her pattern of phobic anxiety and dissociation has led to sexual withdrawal and physical dependency. This may allow an indirect expression of anger, yet also be a compromise between asocial impulses and needs for approval. Unfortunately, this compromise also includes low self-regard and restricted mobility. Another consequence of this pattern is the current low level of available others to provide support. In spite of this, A. G.'s satisfaction with the level of interpersonal support available from her significant other is good and suggests the availability of this individual as a support in any treatment program.

DIAGNOSTIC IMPRESSION

AXIS I	(300.22)	Agoraphobia without recent panic disorder (by history)
	(296.3)	Rule out Major Depressive Disorder, recurrent
		The diagnosis of Anxiety Disorder is based on history rather than current symptomatology. A major differential question has to do with the relevant salience of Major Depression versus Anxiety Disorder
AXIS II		None
AXIS III		Rule out gastrointestinal disorder
AXIS IV		Problems related to social environment-social isolation, restriction of friendships, lifecycle transitions
AXIS V		Current GAF: 62 Highest GAF past year: 80

SUMMARY AND RECOMMENDATIONS

Overall, the foregoing confirms the presence of a complex and long-standing problem founded more in dynamic and early developing interpersonal expectations and conflicts than in symptom-contingent events. However, this is not meant to discount the importance and debilitating nature of the patient's symptom picture, nor the need to give it direct attention in treatment. The dynamic nature of the associated conflicts, and their role in maintaining systemic dysfunction in her relationships with significant others, suggests the need to combine a symptom-focused treatment with efforts to resolve fundamental conflicts. These core conflicts seem to be largely founded in postpubescent strivings to resolve needs for autonomy and dependency. The initial focus of treatment should be on reducing territorial apprehension, with a concomitant increase in social involvement and independent functioning. After initial symptomatic improvement, further interventions should focus on A. G.'s pattern of rebelliousness, which seems to be intertwined with self-incrimination, guilt, and withdrawal. In particular,

this might emphasize confirming needs for both autonomy and acceptance, along with greater insight into this pattern.

A. G.'s level of functional impairment is moderate. Numerous areas of functioning are affected, and this, coupled with the chronicity of the condition, suggests the need for long-term treatment. However, the level of defense and personal control is sufficient, and the level of subjective despair and hopelessness is within a range that suggests that outpatient care is appropriate. There is no evidence of direct risk to self or others. Anxiolytic or antidepressant medications are contraindicated because of her relative degree of control over her symptoms, combined with the high potential for somatic side effects. Individual treatment may allow a more selective and intensive focus on problematic behaviors. Individual therapy would also be likely to prevent the operation of direct avoidance of discomfort when compared with group treatment.

A. G.'s level of distress is well contained since it falls in the average or even below-average range compared with other patients who seek treatment. While her distress increases significantly when exposed to immediate threat, she quickly compensates and is so well versed at avoidance that she may experience difficulty sustaining sufficient motivation for therapy. Thus, interventions that confront or expose her to feared and avoided circumstances may be helpful to desensitize her to anxiety as well as to increase her level of motivation to continue treatment.

A. G.'s coping style vacillates between being primarily impulsive and externalizing, to being self-critical and internalizing. This pattern of cyclic coping suggests the need to address her problems at both a behavioral and an insight level. When her impulses and direct avoidance dominate, behavioral strategies should be emphasized. During phases in which she is more introspective and self-blaming, insight-oriented interventions are likely to be more effective. Given the unsustaining nature of her subjective distress, abreactive and sensate-focused, cathartic interventions may prove to be especially helpful during these more introspective phases.

Finally, A. G. manifests a pattern of superficial compliance and more covert resistance to the directives of help givers. Thus, special attention should be given to developing a trusting relationship. Even if this is achieved, however, she would still be expected to undermine direct suggestions and specific assignments. The most effective approaches, then, would be collaborative interventions emphasizing clear behavioral change, contingency contracting, or paradoxical interventions such as symptom prescription and "no-change" directives. Particular attention may be given to predicting the exacerbation of physical and phobic symptoms following intense sessions because these sessions may mobilize her resistant impulses in an asymptomatic direction.

Collectively, the symptomatic aspects of the patient's fears and phobias may be susceptible to a combination of structured exposure procedures, cognitive restructuring, and interoceptive awareness (Craske & Barlow, 1993). These procedures circumvent patient resistance by virtue of their reliance on self-monitoring, as well as being both symptom and behaviorally focused. The more thematic and dynamic aspects of A. G.'s problem may be addressed by initiating work that specifically mobilizes her anxiety in motivational directions. Confrontation with feared material, along with the use of procedures such as two-chair work and imaginal reliving of unsettling relationships, may be helpful. Imaginal confrontation might be initiated with images and memories of disapproving parents, children, and other significant others, the goals of which may be to

help her tolerate discomfort and disapproval. The procedures outlined by Daldrup, Beutler, Engle, and Greenberg (1988, *Focused Expressive Psychotherapy*) for working with the overcontrolled patient may also be particularly helpful.

RECOMMENDED READING

Armengol, C. G., Kaplan, E., & Moes E. J. (Eds.). (2001). *The consumer-oriented neuropsychological report.* Lutz, FL: Psychological Assessment Resources.

Tallent, N. (1993). *Psychological report writing* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Zuckerman, E. L. (2000). *The clinician's thesaurus: A guidebook for wording psychological reports* (5th ed.). Pittsburgh, PA: Three Wishes Press.

Appendix A

TEST PUBLISHERS/DISTRIBUTORS

American Guidance Service, Inc.
Publisher's Building
Circle Pines, MN 55014-1796
1-800-328-2560
List includes: Kaufman Assessment Battery for Children, Kaufman Adolescent and Adult Intelligence Test, Kaufman Brief Intelligence Test

American Orthopsychiatric Association, Inc.
19 West 44th Street
New York, NY 10036
(212) 354-5770
List includes: Bender Visual-Motor Gestalt Test

Boston Neuropsychological Foundation
P.O. Box 476
Lexington, MA 02173
List includes: Stroop Color-Word Test (Comali/Kaplan version)

Center for Behavioral Health Care Technologies, Inc.
3600 S. Harbor Blvd. #86
Oxnard, CA 93035
(805) 677-4501
www.systematictreatmentselection.com
email: info@cbhti.com
List includes: Systematic Treatment Selection software

Consulting Psychologists Press
3803 East Bayshore Road
P.O. Box 10096
Palo Alto, CA 94303
1-800-624-1765
(650) 969-8901
www.cpp-db.com
www.SkillsOne.com
List includes: California Psychological Inventory, Fundamental Interpersonal Relations Orientation-B, Myers Briggs Type Indicator, Strong Interest Inventory

DLM Teaching Resources
P.O. Box 4000
One DLM Park
Allen, TX 75002
1-800-527-4747
1-800-442-4711
List includes: Woodcock Johnson Psychoeducational Battery

Jastak Associates, Inc.
1526 Gilpin Avenue
Wilmington, DE 19806
1-800-221-WRAT
List includes: Wide Range Achievement Test

Lafayette Instrument Company
P.O. Box 5729
Lafayette, IN 47903
1-800-428-7545
List includes: Hand dynamometer (grip strength)

NFER-Nelson Publishing Co.
DarvilleHouse
2 Oxford Road
East Windsor
Berkshire 21A IDF, UK
List includes: National Adult Reading Test

National Computer Systems, Inc. (NCS)
P.O. Box 1416
Minneapolis, MN 55440
1-800-627-7271
http://assessments.ncspearson.com
List includes: Bender Visual Motor Gestalt Test, Brief Symptom Inventory, Career Assessment Inventory, Children's Depression Inventory, Millon Adolescent Clinical Inventory, Millon Behavioral Health Inventory, Millon Clinical Multiaxial Inventory, Minnesota Multiphasic Personality Inventory, Rorschach, Symptom Checklist-90-R, Sixteen Personality Factor (16 PF), Taylor Johnson Temperament Analysis, Thematic Apperception Test, Test of Memory Malingering

Neuropsychology Laboratory
University of Victoria
P.O. Box 1700
Victoria, BC CANADA
List includes: Paced Auditory Serial Addition Test, Stroop Color-Word Test (Victoria version)

Psychological Assessment Resources
16204 N. Florida Ave.
Lutz, FL 33549
1-800-331-8378
www.parinc.com
List includes: BarOn Emotional Quotient Inventory, Behavioral Assessment of the Dysecutive Syndrome, Bender Visual Motor Gestalt Test, Benton Visual Retention Test, Boston Diagnostic Aphasia Examination, Boston Naming Test, Category Test (computer version), Children's Apperception Test, Children's Auditory Verbal Learning Test, Children's Category Test, Cognitive Assessment System, Color Trails Test, Connor's Rating Scales, Finger Tapping, Hand Dynamometer, Hare Psychopathy Checklist-Revised, House Tree Person, Kaufman Adolescent and Adult Intelligence Test, Kaufman Assessment Battery for Children, NEO-PI-R, Personality Assessment Inventory, Personality Disorder Interview-IV, Repeatable Battery for the Assessment of Neuropsychological Status, Rey Auditory Verbal Learning Test, Rey Complex Figure and Recognition Trial, Rivermead Behavioural Memory Test, Rorschach Self Directed Search, Sentence Completion Series, State Trait Anger Expression Inventory, State Trait Anxiety Inventory, State Trait Anxiety Inventory for Children, Stroop Neuropsychological Screening Test, Stroop Color and Word Test, Tactual Performance Test, Taylor Johnson Temperament Analysis, Test of Everyday Attention, Test of Nonverbal Intelligence, Thematic Apperception Test, Wide Range Achievement Test, Wide Range Assessment of Memory and Learning, Wisconson Card Sorting Test

Psychological Corporation
555 Academic Court
San Antonio, TX 78204
1-800-228-0752
List includes: Beck Depression Inventory, California Verbal Learning Test, Children's Category Test, Children's Memory Scale, Paced Auditory Serial Addition Test, Repeatable Battery for Neuropsychological Status, Rorschach, Wechsler Adult Intelligence Scale, Wechsler Individual Achievement Test, Wechsler Intelligence Scale for Children, Wechsler Preschool and Primary Scale for Children, Wechsler Memory Scale, Wechsler Test of Adult Reading, Wide Range Test of Memory and Learning

Reitan Neuropsychology Laboratory
2920 South 4th Ave.
Tucson, AZ 85713-4819
List includes: Halstead Reitan Neuropsychological Test Battery, Neuropsychological History Questionnaire

Riverside Publishing Co.
8420 Bryn Mawr Avenue
Chicago, IL 60631
(201) 729-6031
List includes: Stanford-Binet

Western Psychological Services
12031 Wilshire Boulevard
Los Angeles, CA 90025-1251
1-800-222-2670
www.wpspublish.com
List includes: AAMR Adaptive Behavior Scales, Adolescent Apperception Test, Behavior Rating Inventory of Executive Functions, Bender Visual Motor Gestalt Test, Children's Category Test, Children's Depression Inventory, Comprehensive Test of Nonverbal Intelligence, Connor's Rating Scales, Draw-a-Person, Eating Disorders Inventory, House-Tree-Person, Human Figure Drawing Test, Family Apperception Test, Kaufman Adolescent and Adult Intelligence Test, Kaufman Assessment Battery for Children, Luria Nebraska Neuropsychological Battery, Millon Index of Personality Styles, Personality Assessment Inventory, Personality Inventory for Children, Psychopathy Checklist, Rey Auditory and Verbal Learning Test, Rivermead Behavioral Memory Test, Rey Auditory Verbal Learning Test, Roberts Apperception Test, Rorschach, Sixteen PF (16 PF), Self-Directed Search, State Trait Anger Expression Inventory, Symbol Digit Modalities Test, Thematic Apperception Test, Taylor Johnson Temperament Analysis, Test of Everyday Attention, Tell Me A Story, Wide Range Achievement Test, Wisconsin Card Sorting

Appendix B

PERCENTILE RANKINGS FOR WECHSLER DEVIATION IQS

IQ	Percentile Rank	IQ	Percentile Rank	IQ	Percentile Rank
155	99.99	118	88	81	10
154	99.98	117	87	80	9
153	99.98	116	86	79	8
152	99.97	115	84	78	7
151	99.97	114	82	77	6
150	99.96	113	81	76	5
149	99.95	112	79	75	5
148	99.93	111	77	74	4
147	99.91	110	75	73	4
146	99.89	109	73	72	3
145	99.87	108	70	71	3
144	99.83	107	68	70	2
143	99.79	106	66	69	2
142	99.74	105	63	68	2
141	99.69	104	61	67	1
140	99.62	103	58	66	1
139	99.53	102	55	65	1
138	99	101	53	64	1
137	99	100	50	63	1
136	99	99	47	62	1
135	99	98	45	61	.47
134	99	97	42	60	.38
133	99	96	39	59	.31
132	98	95	37	58	.26
131	98	94	34	57	.21
130	98	93	32	56	.17
129	97	92	30	55	.13
128	97	91	27	54	.11
127	96	90	27	53	.09
126	96	89	23	52	.07
125	95	88	21	51	.05
124	95	87	19	50	.04
123	94	86	18	49	.03
122	93	85	16	48	.03
121	92	84	14	47	.02
120	91	83	13	46	.02
119	90	82	12	45	.01

Appendix C

CONVERSION FORMULAS AND DIFFERENCE SCORES FOR DETERMINING MAGNITUDE (.05 LEVEL) OF FLUCTUATIONS FOR WECHSLER ADULT INTELLIGENCE SCALE (3RD ED.; WAIS-III) ADDITIONAL GROUPINGS

Additional Grouping	Conversion Formula
Bannatyne's Categories	
Verbal Conceptualization	1.8(V + S + C) + 46
Spatial	1.5(MR + BD + OA + PC) + 40
Sequential	1.6(A + DS + DSymb + L-NS) + 36
Acquired Knowledge	1.9(I + A + V) + 43
Horn Groupings	
Fluid Intelligence	1.1(MR + BD + OA + S + PA + A) + 34
Crystallized Intelligence	1.2(I + V + C + S + PA) + 40
Broad Visualization	1.5(MR + BD + OA + PC) 40
Broad Speediness	1.9(DSymb + SS + OA) + 43
Short-Term Memory	(Same as Working Memory Index)

Appendix D

CONVERSION FORMULAS AND DIFFERENCE SCORES* FOR DETERMINING MAGNITUDE (.05 LEVEL) OF FLUCTUATIONS FOR WECHSLER INTELLIGENCE SCALE FOR CHILDREN (3RD ED.; WISC-III) FACTOR SCORES AND ADDITIONAL GROUPINGS*

Factor or Additional Grouping	Conversion Formula (M = 100, SD = 15)[1]	Difference Score (.05)[2] Mean for All Ages
WISC-III Factors		
Verbal Comprehension	(see Wechsler, 1991, Table A.5,	9.1
Perceptual Organization	A.6, and A.7, pp. 255–257)	10.4
Freedom from Distractibility		11.5
Processing Speed		12
Bannatyne's Categories		
Spatial	2.0(PC+B+OA) + 40	10
Verbal Conceptualization	1.9(V+C+S) + 43	8.5
Sequential	2.3(DSp+A+Coding) + 31	12
Acquired Knowledge	1.9(I+V+A) +43	8.5
SCAD Profile	1.7(SS+C+A+DSp) + 32	9.5
Horn Groupings		
Fluid Intelligence	1.3(S+A+PA+BD+OA) + 34	8.5
Crystallized Intelligence	1.3(I+S+V+C+PA) + 35	9
Achievement	0.85(I+S+A+V+C+PA) + 49	8.5

[1] Age-corrected scores must be used in formulas to calculate standard scores.
[2] Difference scores are calculated by subtracting the standard score on a factor/grouping from the mean of all standard scores within that grouping. The SCAD profile uses the Full Scale IQ as its comparison mean.

*Formulas (excluding the WISC-III factors) are from Kaufman (1994) and difference scores for the factor scores are from Naglieri (1993) and difference scores for the additional groupings were calculated from tables provided by Kaufman, 1994.

Appendix E

DIRECTIONS FOR COMPLETING APPENDIX E: WORKSHEET FOR DETERMINING MAGNITUDE OF WISC-III SUBTEST FLUCTUATIONS

The following directions allow examiners to determine whether subtests fluctuate at a statistically significant level from the full scale, verbal, or performance means. The four steps in this process are numbered to correspond to the numbered (1–4) sections on the worksheet.

1. Decide whether it is appropriate to use either the full scale mean (based on the 10-13 WISC-III subtests) of the verbal and/or performance subtest means. If there is a significant discrepancy between Verbal and Performance IQs (12 points or more to be significant at the .05 level), calculate the mean subtest scores separately for the verbal and performance scales. If Verbal and Performance IQs are not significantly different (e.g., less than 12 points), calculate the mean for all subtests used to develop the Full Scale IQ.
2. Write down the relevant verbal, performance, or full scale mean(s) in the "Mean" column. Calculate the magnitude of any potentially discrepant subtests by noting the differences between age-corrected subtest scores (Step 1) and the relevant means. Record the differences in the "Difference" column.
3. To determine whether the difference is significant, note whether the magnitude of the difference is greater than the value indicated for a specific subtest. For example, if, upon initial appraisal, the WISC-III Information subtest looked as if it might be a significant strength, it would first need to be decided whether the full scale mean or verbal mean would be the most appropriate to use. If the full scale mean was chosen, Information would need to be greater than or equal to the full scale mean by a value of 3.4 points. If the verbal mean was calculated, the difference would need to be greater than or equal to 3. However, be aware that these are average values calculated for all age groups across the standardization samples. Some age groups, particularly among younger populations, may require different values to achieve significance at the set values (see Wechsler, 1991, Table 5.2 for greater precision across different age groups).
4. All subtests that achieved significance should be indicated as either a strength ("S" indicated on the profile sheet) or a weakness ("W"). If a subtest is neither a strength nor a weakness, leave the section blank.

Subtest	1. Subtest Scores			2. Mean	Difference	3. Score required for difference to be significant (.05 level)*			4. Strength or Weakness
	Full Scale	V	P			Full Scale (12 subtests; excluding Mazes)	V	P (excluding Mazes)	
I						3.4	3		
S						3.6	3.1		
A						3.9	3.3		
VB						3	2.7		
C						4	3.4		
DSp						3.3	2.9		
PC						3.9		3.5	
Coding						3.9		3.4	
PA						4		3.6	
BD						3.1		2.9	
OA						4.5		3.9	
SS						4		3.6	
MEAN									

Appendix E.1 Worksheet for determining magnitude (.05) of WISC-III subtest fluctuations

*Data are from Wechsler, 1991.

Appendix F

GUIDELINES FOR HYPOTHESIZING SUBTEST STRENGTHS AND WEAKNESSES

Instructions

To complete Appendixes F1, F2, and F3, you will need to determine values for examinee's subtests based on whether they are *significantly above, above, equal to, below, or signficantly below* their relevant mean score. Which mean score to use (full scale, verbal, performance) and what the means are, have already been determined on the WAIS-III record form or, if using the WISC-III, on Appendix E. For Appendix F1, use the verbal mean if there was a significant difference between Verbal and Performance IQ; otherwise, use the mean for all the subtests administered. Similarly for Appendix F2, use the performance mean if there was a significant difference between Verbal and Performance IQs. For Appendix F3, use verbal means when determining values for verbal subtests and performance means when determining values for performance subtests if there was a significant discrepancy between Verbal and Performance IQs. If there was not a significant difference between Verbal and Performance IQs, use the mean for the total number of subtests administered.

To complete Appendixes F1, F2, and F3, work through the following steps:

1. Designate the following values in the columns directly under the subtests:
   a. *Significantly Above*. Place a "++" in the ability-related boxes in the column(s) under each subtest abbreviation that has been determined to be a significant (.05) strength (see Level IIIa). For example, if someone had a significant strength in Picture Arrangement, all the boxes directly under Picture Arrangement should have ++ placed in them.
   b. *Above.* Place a "+" in the ability-related boxes under each subtest that is greater than 1 scaled score above the relevant mean (but lower than the magnitude required to be significantly above the relevant subtest means).
   c. *Equal to.* Place a "0" in each ability-related box with subtest scores between 1 subtest score above and 1 subtest score below the mean.
   d. *Below.* Place a "–" in each ability-related box with subtest score 1 subscale score below the relevant mean (but not lower than the magnitude required to be significantly above the relevant mean).
   e. *Significantly Below.* Indicate weaknesses by placing a "– –" in the ability-related boxes under each subtest that has been determined to be a significant weakness (see Level IIIa).

2. The next step is to decide whether a hypothesized ability is actually a relative strength (or weakness). The basic strategy is that if a strength (or weakness) is to be accepted, other subtests measuring the same ability should also be high (or low if weaknesses are being determined). However, this is made somewhat more complicated in that there are various numbers of ability boxes in the rows to the right of the ability descriptions. For example, in the first ability listing (Verbal Memory), there are three boxes in the row to the right. Similarly, there are three boxes in the next ability listing (Verbal Conceptualization). However, others have only one box to the right of the ability and others have up to eight. The number of these boxes needs to be taken into account when deciding whether to accept or reject a hypothesized strength or weakness. The following decision rules are recommended:

   One box. In some cases, there are abilities that are specific to a certain subtest and, therefore, have only one ability related to them. If this ability is determined to be a significant strength (++), the hypothesized ability is strengthened. For example, in the ability described as "Practical knowledge and judgment related to conventional standards of behavior," only the Comprehension subtest is indicated as being the subtest that measures this area. Thus, consideration of whether the other subtests were above, equal to, or below the mean is obviously not possible. However, *these subtest specific abilities* should be interpreted with caution with additional outside support provided whenever possible (behavioral observations, relevant history).

   Two boxes. To accept a hypothesized strength composed of a composite of two boxes, one box must be significantly (as determined in Level IIIa) above the mean (++) and the other must also be above the mean (+), although not necessarily significantly above the mean. To accept a hypothesized weakness, the opposite logic would apply in that one box would need to be significantly below the mean (– –) and the other would need to be below the mean (– or – –).

   Three to Four Boxes. To accept strengths composed of composites of three or four subtests, one ability box must be designated as significantly above the relevant mean (++), another must be above the mean (+ or ++), and while it is *preferable* for the third and fourth to be above the mean, one or both are permitted to be at the mean (but none must be below the mean). Again, the opposite logic would be used to accept or reject relative weaknesses.

   Five or More Boxes. To accept a hypothesized strength comprising five or more subtests, at least one of the ability boxes must be a significant strength (++). The rest of the boxes need to be designated as above (+ or ++) or equal (0) to the relevant mean with the exception that one can be below the mean (–), and it is even permissible for it to be significantly below the mean (– –), assuming that most of the other subtests are above the mean. The opposite logic would be used to accept or reject a relative weakness.

3. Examiners should indicate on the far right any strength that has been accepted by writing an "S"; similarly, a "W" should be written if it is a weakness. For example, a person who had a significantly high Picture Arrangement score would have had a "++" placed in the ability-related box for sequencing. If other subtests

measuring sequencing (Arithmetic, Digit Span, Digit Symbol, Letter-Number Sequencing) also had average and/or above-average scores (with one being permissible in the below-average range), this would support the hypothesis that good Sequencing was the (or at least one of the) relative cognitive strength resulting in an elevated score on Picture Arrangement. The higher the corroborating scores from the other subtests also relating to sequencing, the stronger the support that sequencing is the relevant ability. However, note that in most cases abilities will be found to be neither strengths nor weaknesses and thus the box on the far right will need to be left blank.

4. Examiners should work through Steps 1 through 3 only for those abilities in subtests found to be either significantly high or significantly low based on calculations in Level IIIa.

ABILITY	Verbal Subtests							Strength or Weakness
	V	S	A	DSp	I	C	LN-S (WAIS-III)	
Verbal Memory (with Little Verbal Expression)								
Verbal Conceptualization (Concept Formation)								
Fund of Information								
Abstract Verbal Reasoning								
Auditory Short-Term Memory								
Auditory Sequencing								
Verbal Comprehension								
Acquired Knowledge								
Retention								
Language Development and Word Knowledge								
Computational and Numerical Skill								
Practical Knowledge and Judgment related to Conventional Standards of Behavior								
Amount of Benefit from Old Learning or Schooling, Intellectual Curiosity, Range of Interests								
Long-Term Memory								
Complex Verbal Expression								
Simple Verbal Expression								

Appendix F.1 Guidelines for determining subtest strengths and weaknesses (verbal)

ABILITY	Performance								Strength or Weakness
	PC	DSy coding	BD	MR (WAIS-III)	PA	SS	OA	Mazes (WISC-III)	
Visual Organization									
Visual-Motor Coordination									
Visual Perception and Processing of Abstract Information									
Visual Perception and Processing of Meaningful Stimuli									
Nonverbal Reasoning									
Reproduction of Models									
Simutaneous Processing of Visual-Spatial Information									
Visual Sequencing									
Visual Closure									
Visual Memory									
Synthesis									
Trial and Error Learning									
Visual-Spatial Reasoning (Concept Formation)									
Perceptual Organization/Spatial Ability									
Speed of Processing Information									
Planning Ability and Anticipation of Consequences									

Appendix F.2 Guidelines for determining subtest strengths and weaknesses (performance)

ABILITY	Performance								Strength or Weakness
	PC	DSy coding	BD	MR (WAIS-III)	PA	SS	OA	Mazes (WISC-III)	
Analysis of Whole into Component Parts									
Anticipation of Relationships among Parts									
Clerical Speed and Accuracy									
Visual Short-Term Memory									
Spatial Visualization									
Speed of Visual Search									

Appendix F.2 *(Continued)*

ABILITY	Verbal						
	V	S	A	DSp	I	C	LN-S (WAIS-III)
General Ability							
Social Comprehension							
Abstract Reasoning							
Attention and Concentration							
Sequencing							
Fluid Intelligence							
Crystallized Intelligence							
Achievement							
Immediate Role Learning and Recall							
Alertness and Recognition of Relevant from Irrelevant Details							
Alertness to Day-to-Day World							
Ability to Evaluate Past Experience, Social Judgment							
Flexibility of Thinking							
Ability to Evaluate Information							
Short-Term Memory (Visual or Auditory)							

Appendix F.3 Guidelines for determining subtest strengths and weaknesses—using verbal and performance scales combined

PC	DSy Coding	BD	MR	PA	SS	OA	Mazes (WISC-III)	Strength or Weakness

Appendix F.3 *(Continued)*

Appendix G

DIRECTIONS FOR HAND SCORING THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY (MMPI-2) VALIDITY AND CLINICAL SCALES

1. Separate the Scale 5 (*Mf*) scoring keys by sex to correspond with the sex of the person who has taken the test.
2. Items that have been either omitted or double marked should be crossed out with a colored pen, counted as cannot say (?) responses, and the raw score (total number) should be entered on the profile sheet indicated to the right of "? Raw Score."
3. The scoring keys for the validity and clinical scales are then placed over the "softcover answer sheet." The total number of marked items are counted to determine the raw scores for each of the scales. Items marked with a colored pen to designate they are cannot say (?) responses are ignored. The raw scores for each of the scales are entered in the designated sections on the profile sheet. Examiners should make sure that the gender indicated on the profile sheet matches the gender of the examinee.
4. Before plotting the profile, *K* corrections need to be added to the raw scores for *Hs, Pd, Pt, Sc,* and *Ma.* This is done by first calculating the appropriate fractions of *K* (.5*K* to *Hs;* .4*K* to *Pd;* 1*K* to *Pt;* 1*K* to *Sc;* and .2*K* to *Ma*). This can be easily done by using the box to the far left of the profile sheet designated as "Fractions of *K*." The raw score of *K* that was derived from scoring *K* can be located in the far left column of the "Fractions of *K*" box. The three numbers to the right of the raw score of *K* are the correct fractions of *K*. For example, if a raw score for *K* was 15, then .5*K,* .4*K,* and .2*K* would be 8, 6, and 3, respectively. The correct fractions of *K* can then be added to the raw scores for *Hs, Pd,* and *Ma. Pt* and *Sc* both have a full *K* correction added to them.
5. When *K* corrections have been added to *Hs, Pd, Pt, Sc,* and *Ma,* the raw scores can then be plotted on the profile sheet. This can be done by noting the lower scale labels (*L, F, K, Hs*1.5*K,* etc.) and finding the correct raw score on the profile sheet directly above them. These raw scores can then be marked with a dot, circle, or cross. When they have all been marked, a line can be made connecting the 3 validity scales and the 10 clinical scales. The *T* scores can be found by lining up the raw scores with the correct *T* scores on either the far right or the far left of the profile sheet (designated as "T or Tc"). For example, a raw score of 25 on Scale 1 (*Hs*) converts to a *T* score of 80.

Appendix H

FEEDBACK STATEMENTS FOR THE MMPI-2 CLINICAL SCALES

L Scale

The way you approached the test suggests that you were feeling vulnerable to being criticized and judged. You have very high standards and so it is easy for you to feel vulnerable to some kind of moral review. Perhaps you felt that the test would "expose" you in some way and you approached it very cautiously.

F Scale

Your profile suggests that you may be panicked about how much stress you are under right now. You may be so frightened that you are confused about what is happening to you and you want guidance as to what you can do about it.

K Scale

You are a person who is hard to read emotionally. Normally you are not openly nor strongly expressive so that people will have to multiply the intensity of your words to appreciate a sense of empathy for you. For example, if you say you are "somewhat upset" you may in fact be very upset. Sometimes people may mistake your emotional reserve for coldness or indifference.

Scale 1

Your body is a constant source of anxiety and fear for you, so right now your worries about health take up most of your time and energy. So much of your time is spent worrying about your physical well-being, that it is hard for you to accomplish anything, to find outside interests or things you can do that won't incur additional pain. You also may consult a number of physicians for your symptoms without any benefit or relief. Your apprehension and concern about your physical health may leave you feeling somewhat defeated, pessimistic, and bitter. When people try to force you not to focus on your physical problems, you tend to resist them. You feel they don't understand what you are going through. Financial worries, family problems, confrontations, or heavy responsibilities tend to aggravate your physical problems. You may have developed a

Source: From *Therapist Guide to the MMPI & MMPI-2* by R. W. Lewak, P. A. Marks, and G. E. Nelson, 1990, Levittown, PA: Accelerated Development. Reprinted by permission. Conceptual basis of feedback statements derived from Alex Caldwell.

number of ways to get people to help you. At the same time, however, you resent not being able to do things for yourself. People with your profiles often had periods during childhood when they were seriously ill or extremely frightened by the possibility of being physically harmed. Perhaps an explosive or abusive parent or perhaps a serious accident or illness in the family predisposed you to this fear for your physical well being.

When your physical problems and fears seem to get worse, see if, in fact, you're not angry or even afraid of something else. Try to enjoy the times you do feel physically well. Give yourself permission to let go of your fears about your body whenever you feel well.

Scale 2

You are generally a thoughtful, circumspect, and analytical person who takes life and responsibilities very seriously. However, right now some of your strengths may be working against you, for you are feeling very dejected, gloomy, and depressed. You are probably spending most of your time thinking about things you've done or said and feeling hopeless about things ever changing. While you may be able to do some basic chores, you may find little motivation to pursue enjoyable activities. In fact, low energy is probably a cause of concern to you. Unfortunately, you probably blame yourself for your loss of energy and lack of interest. This compounds your sense of hopelessness, helplessness, and despair. Your sleep may be constantly interrupted so you awaken feeling tired and without energy. Your appetite may be markedly diminished, or you may take comfort in eating. Your interest in sex may be low. You may find it difficult to concentrate, make decisions, and plan for the future.

Try to find small rewards and pleasures that you can give yourself on a daily basis and keep a record of these. When you feel pessimistic about the future, force yourself (if necessary) to write down some of the things that have gone well for you recently, so you can keep things in perspective. Try to stop from constantly blaming yourself for things you think have gone wrong. Begin a regular exercise program. Keep a record of your accomplishments and things that you have done well.

Scale 3

You tend to be an agreeable, even sentimental and romantic person who yearns for a life where people are kind and loving to each other. Very likely you work hard avoiding interpersonal difficulties or holding controversial views. You probably seek out situations where you can please others, make them happy and relieve them of suffering. Your discomfort with anger probably makes it difficult for you to confront people. People with your profile are often seen as playful and childlike because of their extreme discomfort with the adult world of competition, greed, and cruelty. Serious and painful responsibilities are something you avoid. You are easily influenced by other people's philosophies and think that you should think, want, and do what they do. Holding your feelings in and trying to stay positive, however, is putting stress on your body and you may be having all or any of the following symptoms: headaches, backaches, neckaches, stomach aches, nausea, or dizziness. These symptoms probably get worse as your stress increases and may change depending on what kind of stress you are under. Rejection and anger from

others are extremely painful for people with your profile. Perhaps as a child one of your parents was explosive or abusive and frightened you. Your reaction very likely was to try to be brave, and to look for a positive outcome. You may have learned to be totally unaware of your anger. Headaches, stomachaches, low back pain, or other physical symptoms are often caused by unfelt and unexpressed angry and negative feelings.

When you experience physical symptoms, such as headaches, stomachaches, etc., look to see if you are struggling with some angry feelings which are difficult to express. Whenever you find yourself even mildly resentful or angry towards someone, try to express your feelings to them immediately even in small matters. Try to see what is negative, as well as positive, in any given situation and try to balance the two extremes.

Scale 4

People with your profile are very independent and rather uncomfortable working for others. Right now it is difficult for you to care about others and you don't want to get involved with them for fear they will disappoint you or let you down. You want to protect yourself –you don't want to count on anyone again and so you will keep your distance and stay uninvolved and emotionally numb. That way, you hope to avoid a repetition of previous pain. You are probably a survivor. You learned to take care of yourself at an early age. This may have left you with a cynical view of the world in which real love and caring doesn't exist. You may also feel angry, bitter, and defeated in your relationships. You have a low tolerance for frustration and you may find yourself restless and impulsive, especially when stressed. You like immediate gratification. You want what you want, and what it right now. In some cases this can lead to problems with the law and to problems with authority as well. People with your profile often have numerous relationships with the opposite sex, but there is little real intimacy or satisfaction. In fact, letting yourself really care about someone is difficult because you do not expect people to care about you. Often they will view the world as a cruel and heartless place. They fear that if they are not vigilant and should become vulnerable, others will use them and take advantage of them. People with your profile often grew up in environments where they had to care for themselves because the authority figures in their lives could not be trusted to meet their needs. You may remember specific occasions in your childhood when you were particularly disappointed.

When you get bored and look for excitement, find things to do that are not dangerous, illegal, or destructive. When you set a goal, try to give yourself small rewards along the way so that you don't think about quitting before you have given it a real try. Find ways to express your anger before it starts to build instead of letting it accumulate and explode.

Scale 6

Right now you are feeling extremely sensitive to criticism, attack, or judgment. You are also very cautious about revealing to others your deepest thoughts and feelings. This is probably because you fear you may reveal too much and your self-revelations will be used against you. You are wary of being a victim of someone else's power. This makes you very cautious and concerned about the motives of others. If you care about someone deeply, it frightens you because you feel vulnerable to them. You feel that this

gives them control over you and then they may hurt you. Being rational and fair is also extremely important to you, and you expect others to be this was as well. There are times when you become so angry that it is hard for you to forgive the person who you see as the source of your anger and who has frustrated you. People with your profile often were reared in households where their parents were controlling and judgmental. As a child you may have felt unfairly or unjustly criticized and attacked. At present you may experience some fears of being criticized which makes you feel very tense and cautious. It reawakens your vulnerability to being dominated and controlled.

At times when you feel you are being criticized unfairly, remember that not all criticism is intended as a personal attack on you. If you find yourself feeling mildly criticized by someone, observe to see if your feelings get hurt and you withdraw rather than standing up to argue your case.

Scale 7

Right now you are feeling constantly on the alert as though danger lurked around every corner. It is hard for you to think clearly, remember things, or concentrate, because you are constantly distracted and fretful about details. You may find yourself making "mental lists" of all the things you have to worry about to avoid some unpredictable catastrophe. People with your profile often develop a number of habits or superstitions that serve to reduce their anxiety. You tend to be a careful, thorough, and persevering person who attends to every detail. Confrontations with others that might lead to anger are frightening and you tend to avoid them. Being angry with others causes you to feel guilty, tense, and uncomfortable. Consequently, being assertive is generally difficult for you and you may allow others to "push you around" or take advantage of you. People with your profile often were reared in homes where they were subjected to unreasonable or unexpected and frightening events. They tried to protect themselves from the unexpected by analyzing and predicting the future and thereby avoiding any future painful event.

Whenever you find yourself tense with your mind vigilantly attentive, take some time and think of a peaceful and relaxing scene. Once you make a decision, do not go back over it and reanalyze it. When something does go wrong, try not to punish and blame yourself as a bad person.

Scale 8

You are probably feeling very separate from others and it is hard for you to know how others feel towards you. You are probably confused about your thoughts and moods and wonder if there is something really wrong with you. Perhaps you feel that if people get too close to you and come to know you, they will somehow discover that something is wrong with you and reject you. Very likely you spend a good deal of time day-dreaming, and it is hard for you to organize your thoughts or make decisions and get things accomplished. Often odd and unpredictable things will intrude into your thinking that are uncomfortable and even terrifying. You may feel very lonely but at the same time you desperately avoid getting close to people and allowing them to get close to you. If anyone is hostile or angry toward you, it completely disorganizes you making it hard for you to think clearly and respond. There are times when you may say something you think is appropriate, funny, or

sympathetic to someone and they will respond in an angry or indifferent way. People with your profile often grew up in homes where they were subjected to a great deal of anger and hostility from the important adults in their lives. You may have protected yourself by daydreaming and trying to escape into fantasies.

If you find yourself in a sad or bad mood and feel alienated from others, rather than withdrawing into the mood, force yourself to find something to do that could make you feel better. Talk yourself out of withdrawing as a solution. When you feel a bad mood coming on, avoid scolding yourself and feeling that you are a terrible person. Although it may seem difficult, focus on thoughts that make you feel good, even if you must force yourself to do so.

Scale 9

You are a person with a very high level of energy and you are unusually optimistic, even in situations where other people might feel defeated or disgruntled. You may find yourself so optimistic that you make promises that are hard to keep. Working on one task is difficult for you. You get very excited by the prospect of something new and then it is hard for you to rest or to focus on the task at hand. You need novelty, excitement, and challenge. You may find yourself impatient with others. The world appears to be moving too slowly for you.

Make a list of things you wish to do in the immediate future and start to do them, one by one, without being distracted. Try to finish one thing before beginning another. Reward yourself frequently and regularly for completing tasks. Try not to commit yourself to too many task and activities. Do not make any changes in your career or life goals until you have discussed these and thought them through with someone else.

Appendix I

DESCRIPTIVE STATISTICS FOR NONPATIENT ADULTS (N = 600)

Variable	Mean	*SD*	Min	Max	Freq	Median	Mode	SK	KU
R	22.32	4.40	14.00	43.00	600	22.00	23.00	0.86	1.90
W	8.28	2.36	3.00	24.00	600	8.00	9.00	1.67	7.82
D	12.88	3.77	0.00	32.00	598	13.00	14.00	–0.14	1.72
Dd	1.16	[1.67]	0.00	15.00	370	1.00	0.00	4.00	24.01
S	1.57	[1.28]	0.00	10.00	514	1.00	1.00	1.99	7.61
DQ+	7.36	2.23	1.00	19.00	600	7.00	6.00	0.53	1.24
DQo	13.58	3.67	5.00	36.00	600	14.00	15.00	1.26	5.69
DQv	0.98	[1.26]	0.00	6.00	306	1.00	0.00	1.35	1.30
DQv/+	0.39	[0.61]	0.00	2.00	193	0.00	0.00	1.32	0.65
FQx+	0.71	[0.88]	0.00	5.00	290	0.00	0.00	1.33	2.19
FQxo	16.44	3.34	7.00	29.00	600	17.00	17.00	0.25	0.59
FQxu	3.49	2.03	0.00	16.00	580	3.00	3.00	1.50	5.33
FQx-	1.56	1.20	0.00	8.00	513	1.00	1.00	1.25	2.58
FQxNone	0.11	[0.37]	0.00	3.00	60	0.00	0.00	3.80	17.53
MQ+	0.44	[0.68]	0.00	3.00	210	0.00	0.00	1.52	1.98
MQo	3.57	1.84	0.00	8.00	595	3.00	3.00	0.42	–0.62
MQu	0.21	0.51	0.00	5.00	104	0.00	0.00	3.24	16.14
MQ-	0.07	[0.27]	0.00	2.00	35	0.00	0.00	4.48	21.40
MQNone	0.01	[0.08	0.00	1.00	4	0.00	0.00	12.15	146.23
S-	0.25	[0.56]	0.00	3.00	117	0.00	0.00	2.71	8.25
M	4.30	1.95	1.00	10.00	600	4.00	3.00	0.48	–0.55
FM	3.74	1.31	0.00	9.00	598	4.00	4.00	0.15	0.58
m	1.28	0.99	0.00	6.00	458	1.00	1.00	0.62	0.61
FM+m	5.01	1.70	0.00	12.00	5.99	5.00	5.00	0.20	0.25
FC	3.56	1.88	0.00	9.00	580	3.00	3.00	0.38	–0.24
CF	2.41	1.31	0.00	7.00	564	2.00	3.00	0.29	–0.17
C	0.12	[0.37]	0.00	3.00	61	0.00	0.00	3.76	17.14
Cn	0.01	[0.08]	0.00	1.00	4	0.00	0.00	12.15	146.23
Sum Color	6.09	2.44	0.00	12.00	599	6.00	5.00	0.11	–0.66
WSumC	4.36	1.78	0.00	9.50	599	4.00	3.50	0.11	–0.54
Sum C′	1.49	[1.16]	0.00	10.00	490	1.00	1.00	1.41	5.96

Note: Standard deviations shown in brackets indicate that the value is probably unreliable and/or misleading and should not be used to estimate expected ranges. Ordinarily these variables should not be included in most parametric analyses.

Source: From *The Rorschach: A Comprehensive System I: Basic Foundations* (2nd ed.), pp. 260–262, by J. E. Exner Jr. Copyright © 1993 by John E. Exner Jr. Reprinted by permission of John Wiley & Sons.

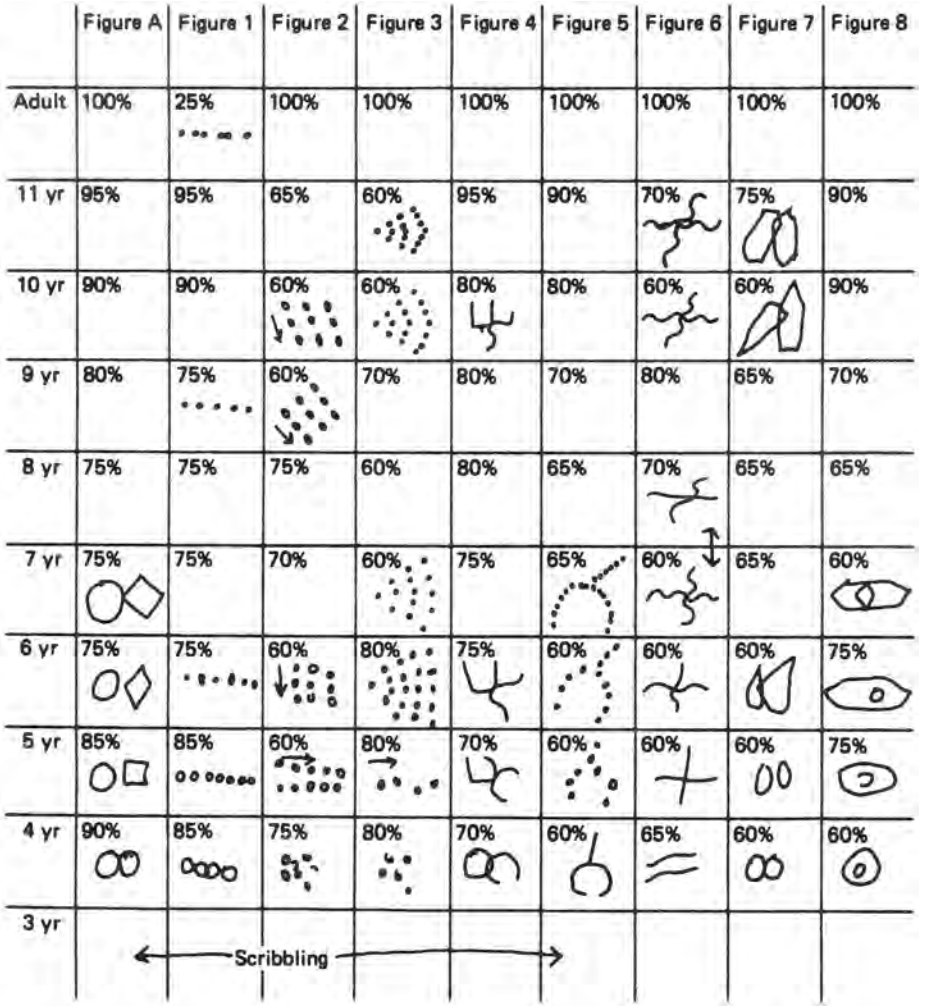
Variable	Mean	*SD*	Min	Max	Freq	Median	Mode	SK	KU
Sum T	0.95	[0.61]	0.00	4.00	490	1.00	1.00	0.83	3.33
Sum V	0.28	[0.61]	0.00	5.00	124	0.00	0.00	2.71	9.58
Sum Y	0.61	[0.96]	0.00	10.00	262	0.00	0.00	3.53	23.46
Sum Shading	3.32	2.09	0.00	23.00	588	3.00	3.00	2.54	15.45
Fr + rF	0.11	[0.43]	0.00	4.00	48	0.00	0.00	4.98	
30.45*FD*	1.18	[0.94]	0.00	5.00	456	1.00	1.00	0.84	1.35
F	7.95	2.83	2.00	23.00	600	8.00	7.00	0.92	2.04
(2)	8.52	2.18	1.00	21.00	600	8.00	8.00	0.29	2.11
3r + (2)/R	0.40	0.09	0.03	0.87	600	0.39	0.33	0.47	3.86
Lambda	0.60	0.31	0.11	2.33	600	0.53	0.50	2.27	8.01
EA	8.66	2.38	2.00	18.00	600	9.00	9.50	–0.04	0.42
es	8.34	2.99	3.00	31.00	600	8.00	7.00	1.43	6.58
D Score	–0.03	0.97	–10.00	3.00	600	0.00	0.00	–3.06	24.34
AdjD	0.15	0.82	–5.00	3.00	600	0.00	0.00	–0.88	5.89
a (active)	6.44	2.23	0.00	14.00	599	6.00	6.00	0.32	0.01
p (passive)	2.90	1.64	0.00	9.00	572	3.00	2.00	0.57	0.03
Ma	2.90	1.57	0.00	8.00	583	3.00	2.00	0.52	–0.26
Mp	1.42	1.03	0.00	5.00	493	1.00	1.00	0.53	–0.13
Intellect	1.57	1.48	0.00	9.00	449	1.00	1.00	1.27	2.16
Zf	11.84	2.78	5.00	27.00	600	12.00	12.00	0.87	3.44
Zd	0.57	2.98	–11.50	9.50	560	0.50	–1.00	0.31	0.48
Blends	5.15	2.08	0.00	12.00	598	5.00	5.00	0.00	–0.26
Blends/R	0.24	0.10	0.00	0.67	598	0.24	0.26	0.35	0.65
Col-Shd Blends	0.45	[0.68]	0.00	5.00	215	0.00	0.00	1.70	4.12
Afr	0.67	0.16	0.23	1.29	600	0.67	0.67	0.35	0.65
Populars	6.58	1.39	3.00	10.00	600	6.00	6.00	–0.09	–0.47
XA%	0.92	0.06	0.57	1.00	600	0.94	0.96	–1.34	3.68
WDA%	0.94	0.06	0.54	1.00	600	0.95	1.00	–1.42	4.93
X+%	0.77	0.09	0.35	1.00	600	0.78	0.80	–0.86	2.33
X-%	0.07	0.05	0.00	0.43	513	0.05	0.04	1.41	4.56
X*u*%	0.15	0.07	0.00	0.45	580	0.15	0.13	0.54	0.86
Isolate/R	0.19	0.09	0.00	0.60	588	0.18	0.16	0.51	0.41
H	3.21	1.71	0.00	9.00	595	3.00	2.00	0.97	0.84
(H)	1.22	1.02	0.00	6.00	432	1.00	1.00	0.65	0.48
Hd	0.84	[1.02]	0.00	7.00	336	1.00	0.00	1.98	6.60
(Hd)	0.21	[0.50]	0.00	4.00	109	0.00	0.00	2.90	11.25
Hx	0.03	[0.23]	0.00	4.00	14	0.00	0.00	11.29	164.54
All *H* Cont	5.49	1.75	1.00	15.00	600	5.00	5.00	0.59	1.24
A	7.96	2.04	3.00	25.00	600	8.00	7.00	1.06	5.03
(A)	0.27	[0.47]	0.00	3.00	137	0.00	0.00	2.31	6.38
Ad	2.30	[1.18]	0.00	9.00	571	2.00	2.00	0.79	2.85
(Ad)	0.10	[0.26]	0.00	2.00	53	0.00	0.00	3.57	13.07
An	0.54	[0.77]	0.00	4.00	243	0.00	0.00	1.59	2.81
Art	0.90	0.91	0.00	5.00	363	1.00	0.00	0.98	1.20
Ay	0.35	[0.52]	0.00	3.00	198	0.00	0.00	1.23	1.38

(continued)

Variable	Mean	*SD*	Min	Max	Freq	Median	Mode	SK	KU
Bl	0.20	[0.46]	0.00	3.00	104	0.00	0.00	2.40	5.80
Bt	2.37	1.32	0.00	6.00	551	2.00	3.00	0.17	–0.29
Cg	1.41	1.09	0.00	5.00	482	1.00	1.00	0.73	0.29
Cl	0.14	[0.38]	0.00	2.00	78	0.00	0.00	2.67	6.76
Ex	0.20	[0.40]	0.00	2.00	119	0.00	0.00	1.57	0.74
Fi	0.56	[0.77]	0.00	4.00	240	0.00	0.00	1.09	0.22
Food	0.21	[0.47]	0.00	3.00	112	0.00	0.00	2.26	5.03
Ge	0.05	[0.24]	0.00	2.00	27	0.00	0.00	5.18	28.97
Hh	0.99	0.90	0.00	4.00	407	1.00	1.00	0.85	0.57
Ls	0.86	0.79	0.00	3.00	382	1.00	1.00	0.60	–0.23
Na	0.36	[0.63]	0.00	6.00	178	0.00	0.00	2.35	11.12
Sc	1.12	[1.15]	0.00	6.00	388	1.00	0.00	1.22	1.96
Sx	0.11	[0.47]	0.00	5.00	46	0.00	0.00	6.16	48.09
Xy	0.05	[0.24]	0.00	2.00	29	0.00	0.00	4.80	24.46
Idio	1.36	[1.32]	0.00	7.00	404	1.00	0.00	1.03	1.43
DV	0.59	[0.78]	0.00	4.00	266	0.00	0.00	1.36	1.77
INCOM	0.56	[0.78]	0.00	4.00	263	0.00	0.00	1.74	3.91
DR	0.39	[0.69]	0.00	4.00	175	0.00	0.00	1.97	4.15
FABCOM	0.27	[0.52]	0.00	3.00	141	0.00	0.00	1.85	3.02
DV2	0.00	[0.06]	0.00	1.00	2	0.00	0.00	17.27	297.49
INC2	0.02	[0.13]	0.00	1.00	10	0.00	0.00	7.57	55.49
DR2	0.01	[0.11]	0.00	1.00	8	0.00	0.00	8.50	70.61
FAB2	0.03	[0.16]	0.00	1.00	16	0.00	0.00	5.89	32.81
ALOG	0.04	[0.20]	0.00	2.00	21	0.00	0.00	5.58	33.07
CONTAM	0.00	[0.00]	0.00	0.00	0	0.00	0.00	—	—
Sum6 Sp Sc	1.91	1.47	0.00	7.00	496	2.00	1.00	0.80	0.56
Lvl 2 Sp Sc	0.06	[0.25]	0.00	2.00	34	0.00	0.00	4.33	19.52
WSum6	4.48	4.08	0.00	28.00	496	4.00	0.00	1.42	3.25
AB	0.16	[0.43]	0.00	3.00	84	0.00	0.00	2.82	8.39
AG	1.11	1.15	0.00	5.00	380	1.00	0.00	1.02	0.60
COP	2.00	1.38	0.00	6.00	498	2.00	2.00	0.25	–0.63
CP	0.01	[0.09]	0.00	1.00	5	0.00	0.00	10.84	115.98
GOODHR	4.93	1.78	0.00	10.00	598	5.00	5.00	0.36	0.02
POORHR	1.53	1.46	0.00	8.00	431	1.00	1.00	1.25	2.30
MOR	0.79	[0.89]	0.00	4.00	321	1.00	0.00	1.01	0.60
PER	0.92	[0.91]	0.00	5.00	385	1.00	1.00	1.33	3.39
PSV	0.07	[0.25]	0.00	2.00	38	0.00	0.00	3.84	14.28

Appendix J

MATURATIONAL GUIDELINES FOR BENDER GESTALT DESIGNS



Copyright © 1938. The Riverside Publishing Company. All rights reserved. Reproduced from *A Visual Motor Gestalt Test and Its Clinical Uses,* by Lauretta Bender with permission of the publisher.

Appendix K

REY AUDITORY-VERBAL LEARNING TEST SCORING SHEET

RAVLT Scoring Sheet

Name _____

Date _____

Examiner _____

(Note: Do not re-read List A for Recall Trial 7 or 8)

List A	Recall Trials					List B	Recall Trials				
	1	2	3	4	5		6	7	8	9*	
drum						desk					drum
curtain						ranger					curtain
bell						bird					bell
coffee						shoe					coffee
school						stove					school
parent						mountain					parent
moon						glasses					moon
garden						towel					garden
hat						cloud					hat
farmer						boat					farmer
nose						lamb					nose
turkey						gun					turkey
color						pencil					color
house						church					house
river						fish					river
#correct:											

Scoring Categories:

1. Immediate Memory (score for Trial 1) _____
2. Best Learning (score for Trial 5) _____
3. Total Learning (sum of the scores for Trials 1–5) _____
4. Proactive Interference (score for Trial 6) _____
5. Retroactive Interference (score for Trial 7) _____
6. Delayed Recall (score for Trial 8) _____
7. Recognition (score for Trial 9) _____

* Trial 9: Determine the number of words correctly/incorrectly recognised from the following recognition list.

bell (A)	home (SA)	towel (B)	boat (B)	glasses (B)
window (SA)	fish (B)	curtain (A)	hot (PA)	stocking (SB)
hat (A)	moon (A)	flower (SA)	parent (A)	shoe (B)
barn (SA)	tree (PA)	color (A)	water (SA)	teacher (SA)
ranger (B)	balloon (PA)	desk (B)	farmer (A)	stove (B)
nose (A)	bird (B)	gun (B)	rose (SPA)	nest (SPB)
weather (SB)	mountain (B)	crayon (SA)	cloud (B)	children (SA)
school (A)	coffee (A)	church (B)	house (A)	drum (A)
hand (PA)	mouse (PA)	turkey (A)	stranger (PB)	toffee (PA)
pencil (B)	river (A)	fountain (PB)	garden (A)	lamb (B)

Note: (A) words from list A; (B) words from list B; (S) word with a semantic association to a word on list A or B as indicated; (P) word phonemically similar to a word on list A or B.

Adapted from: Lezak (1995) and Spreen & Strauss (1991).

Appendix L

NORMS FOR THE STANDARD ADMINISTRATION OF THE RAVLT

	Trial																			
	1		2		3		4		5		Sum 1–5		6		7		Delayed Recall		Recog-nition	
Age	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
13	6.8	1.9	9.4	2.1	11.4	1.9	12.1	1.7	13.1	1.3	52.8	7.0	6.2	2.1	11.8	2.0	12.0	2.1	14.1	1.3
14–15	6.7	1.6	9.5	1.8	11.4	1.9	12.2	1.9	12.9	1.7	52.7	7.0	6.1	1.6	11.6	2.2	11.8	2.4	14.4	1.0
16–19	6.8	1.6	9.2	2.0	11.4	1.7	12.3	1.4	12.8	1.4	53.9	6.7	6.5	1.7	11.4	2.4	11.7	2.2	14.2	1.2
20–29	7.0	1.8	9.9	1.8	11.5	2.1	12.4	1.9	12.9	1.8	56.1	7.3	6.7	2.0	11.5	2.3	11.3	2.3	14.3	1.1
30–39	6.7	1.8	9.9	2.2	11.4	2.2	12.2	2.0	12.7	1.9	53.6	8.3	6.5	2.0	11.2	2.7	11.1	2.8	14.2	1.2
40–49	6.6	1.7	9.3	1.9	10.8	2.1	11.7	2.1	12.3	1.9	51.1	8.6	6.1	1.9	10.4	2.8	10.2	2.8	14.0	1.4
50–59	6.2	1.6	9.0	1.9	10.5	1.9	11.4	1.9	12.1	2.1	47.6	8.1	5.7	2.2	9.9	2.8	9.9	3.2	13.9	1.4
60–69	5.9	1.6	8.4	2.0	9.8	2.3	10.9	2.3	11.3	2.3	43.4	7.7	5.1	1.3	9.3	2.9	8.8	3.0	13.5	1.3
70–79	5.5	1.6	7.7	2.1	8.8	2.1	9.8	2.4	10.3	2.4	37.1	7.5	3.9	1.6	8.1	3.0	7.0	2.4	13.3	1.5

Note: Adapted from Schmidt, 1996.

References

Abrams, E. (1955). Prediction of intelligence from certain Rorschach factors. *Journal of Clinical Psychology, 11,* 81–83.

Achenbach, T. M. (1994). *Child Behavior Checklist.* Burlington, VT: Author.

Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile.* Burlington: University of Vermont, Department of Psychiatry.

Acker, M. B. (1990). A review of the ecological validity of neuropsychological tests. In D. E. Tupper & K. D. Cicerone (Eds.), *The neuropsychology of everyday life: Assessment and basic competencies* (pp. 19–55). Boston: Kluwer Academic.

Acker, M. B., & Davis, J. R. (1989). Psychology test scores associated with late outcome head injury. *Neuropsychology, 3,* 1–10.

Ackerman, P. T., Peters, J. R., & Dykman, R. A. (1971). Children with specific learning disabilities: Bender Gestalt Test findings and other signs. *Journal of Learning Disabilities, 4,* 437–446.

Ackerman, S. J., Clemence, A. J., Weatherill, R., & Hilsenroth, M. J. (1999). Use of the TAT in the assessment of *DSM-IV* personality disorders. *Journal of Personality Assessment, 73,* 422–448.

Ackerman, S. J., Hilsenroth, M. J., Baity, M. R., & Blagys, M. D. (2000). Interaction of therapeutic process and alliance during psychological assessment. *Journal of Personality Assessment, 75,* 82–109.

Acklin, M. W., & Bernat, E. (1987). Depression, alexithymia, and pain prone disorder: A Rorschach study. *Journal of Personality Assessment, 51,* 462–479.

Acklin, M. W., McDowell, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *Journal of Personality Assessment, 74,* 15–47.

Adams, H., Cooper, G., & Carrera, R. (1963). The Rorschach and the MMPI: A concurrent validity study. *Journal of Projective Techniques, 27,* 23–24.

Adams, K. M. (1980). In search of Luria's battery: A false start. *Journal of Consulting and Clinical Psychology, 48,* 511–516.

Adams, W., & Sheslow, D. (1990). *WRAML manual.* Wilmington, DE: Jastak Associates.

Adcock, C. J. (1965). Thematic Apperception Test: A review. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (Vol. 1, pp. 533–535). Highland Park, NJ: Gryphon Press.

Affleck, D. C., & Mednick, S. A. (1959). The use of the Rorschach Test in the prediction of the abrupt terminator in individual psychotherapy. *Journal of Consulting Psychology, 23,* 125–128.

Ahava, G. W., Iannone, C., Grebstein, L., & Schirling, J. (1998). Is the Beck Depression Inventory reliable over time? An evaluation of multiple test-retest reliability in a nonclinical college student sample. *Journal of Personality Assessment, 70,* 222–231.

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *Journal of Counseling Psychology, 48,* 251–257.

Albert, S., Fox, H. M., & Kahn, M. W. (1980). Faking psychosis on the Rorschach: Can expert judges detect malingering? *Journal of Personality Assessment, 44,* 115–119.

Alcock, T. (1963). *The Rorschach in practice.* Philadelphia: Lippincott.

Alessi, G. J. (1980). Behavioral observation for the school psychologist: Responsive-discrepancy model. *School Psychology Review, 9,* 31–45.

Alfonso, V. C., Johnson, A., Patinella, L., & Radar, D. E. (1998). Common WISC-III examiner errors: Evidence from graduate students in training. *Psychology in the Schools, 35,* 119–125.

Allen, J. J., Iacono, W. G., & Danielson, K. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology, 29,* 504–522.

Allen, R. M. (1954). *Elements of Rorschach interpretation.* New York: Harper & Row.

Allison, J., & Blatt, S. J. (1964). The relationship of Rorschach whole responses to intelligence. *Journal of Projective Techniques, 28,* 255–260.

Allison, J., Blatt, S. J., & Zimet, C. N. (1968). *The interpretation of psychological tests.* New York: Harper & Row.

Alloy, L. B., Abramson, W. G., Whitehouse, W. G., Hogan, M. E., Tashman, N. A., Steinberg, D. L., et al. (1999). Depressogenic cognitive styles: Predictive validity, information processing, and personality characteristics, and developmental origins. *Behavior Research and Therapy, 37,* 503–531.

Alnacs, R., & Torgerson, S. (1989). Clinical differentiation between major depression only, major depression with panic disorder, and panic disorder: Childhood personality, and personality disorder. *Acta Psychiatrica Scandinavica, 79,* 370–377.

Amabile, T. M. (1983). *The social psychology of creativity.* New York: Springer-Verlag.

Ambrosini, P. J. (2000). Historical development and present status of the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS). *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 49–58.

Ambrosini, P. J., Metz, C., Prabucki, K., & Lee, J. (1989). Videotape reliability of the third revised edition of K-SADS. *Journal of the American Academy of Child and Adolescent Psychology, 28,* 723–728.

American Educational Research Association. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

American Journal of Managed Care. (1999). Introduction. *American Journal of Managed Care, 5,* S764–S766.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychological Association. (1986). *Guidelines for computer-based test interpretations.* Washington, DC: Author.

American Psychological Association. (1987). General guidelines for providers of psychological services. *American Psychologist, 42,* 7.

American Psychological Association. (1988). *Computer use in psychology.* Washington, DC: Author.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47,* 1597–1611.

American Psychological Association. (1994). Guidelines for child custody evaluations in divorce proceedings. *American Psychologist, 49,* 677–680.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

American Psychological Association Committee on Professional Practice and Standards. (1998). *Guidelines for psycholotgical evaluations in child protection matters.* Washington, DC: American Psychological Association.

Ames, L. B. (1959). Further check on the diagnostic validity of the Ames danger signals. *Journal of Projective Techniques, 23,* 291–298.

Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. N. (1973). *Rorschach responses in old age* (Rev. ed.). New York: Brunner/Mazel.

Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. N. (1974). *Child Rorschach responses: Developmental trends from two to ten years* (Rev. ed.). New York: Brunner/Mazel.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Anastopoulos, A. D., Spisto, M. A., & Maher, M. C. (1994). The WISC-III Freedom from Distractibility Factor: Its utility in identifying children with Attention Deficit Hyperactive disorder. *Psychological Assessment, 6,* 368–371.

Anderson, E. M., & Lambert, M. J. (1995). Short-term dynamically oriented psychotherapy: A review and meta-analysis. *Clinical Psychology Review, 15,* 503–514.

Anderson, T. K., Cancelli, A. A., & Kratochwill, T. R. (1984). Self-reported assessment practices of school psychologists: Implications for training and practice. *Journal of School Psychology, 22,* 17–29.

Andrews, H. B. (2001). Back to basics: Psychotherapy is an interpersonal process. *Australian Psychologist, 36,* 107–114.

Angus, L. E., Levitt, H., & Hardtke, K. (1999). The narrative process coding system: Research applications and implications for psychotherapy. *Journal of Clinical Psychology, 55,* 1255–1270.

Antonuccio, D. O., Danton, W. G., & DeNelsky, G. Y. (1995). Psychotherapy versus medication for depression: Challenging the conventional wisdom with data. *Professional Psychology: Research and Practice, 26,* 574–585.

Antony, M. M., & Barlow, D. (Eds.). (2002). *Handbook of assessment, treatment planning, and outcome evaluation: Empirically supported strategies for psychological disorders.* New York: Guilford Press.

Apter, A., Bleich, A., Plutchik, R., Mendelsohn, S., & Tyrano, S. (1988). Suicidal behavior, depression, and conduct disorder in hospitalized adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 696–699.

Aram, D. M., & Ekelman, B. L. (1986). Cognitive profiles of children with early onset unilateral lesions. *Developmental Neuropsychology, 2,* 155–172.

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency Psychopathology Scale, F(p). *Psychological Assessment, 7,* 424–431.

Arbisi, P. A., & Ben-Porath, Y. S. (1998). Characteristics of the MMPI-2 F(p) Scale as a function of diagnosis in an inpatient VA sample. *Psychological Assessment, 10,* 221–228.

Archer, R. P. (1984). Use of the MMPI with adolescents: A review of salient issues. *Clinical Psychology Review, 4,* 241–251.

Archer, R. P. (1987). *Using the MMPI with adolescents.* Hillsdale, NJ: Erlbaum.

Archer, R. P. (1992a). *MMPI-A: Assessing adolescent psychopathology.* Hillsdale, NJ: Erlbaum.

Archer, R. P. (1992b). Review of the Minnesota Multiphasic Personality Inventory-2. In J. J. Kramer & J. C. Conoley (Eds.), *Eleventh mental measurements yearbook* (pp. 558–562). Lincoln, NE: Buros Institute of Mental Measurements.

Archer, R. P., & Gordon, R. (1988). MMPI and Rorschach indices of schizophrenic and depressive diagnosis among adolescent inpatients. *Journal of Personality Assessment, 52,* 276–287.

Archer, R. P., Griffin, R., & Aiduk, R. (1995). MMPI-2 clinical correlates for ten common codes. *Journal of Personality Assessment, 65,* 391–407.

Archer, R. P., & Jacobson, J. M. (1993). Are critical items "critical" for the MMPI-A? *Journal of Personality Assessment, 61,* 547–556.

Archer, R. P., & Krishnamurthy, R. (1997a). MMPI-A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment, 69,* 517–533.

Archer, R. P., & Krishnamurthy, R. (1997b). MMPI-A scale-level factor structure: Replication in a clinical sample. *Assessment, 4,* 337–349.

Archer, R. P., & Krishnamurthy, R. (2003). The Rorschach. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (2nd ed.). New York: Guilford Press.

Archer, R. P., Maruish, M. E., Imhof, E. A., & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. *Professional Psychology: Research and Practice, 22,* 247–252.

Archer, R. P., & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment, 7,* 227–236.

Archer, R. P., Pancoast, D. L., & Klinefelter, D. (1989). A comparison of MMPI code types produced by traditional and recent adolescent norms. *Psychological Assessment, 1,* 23–29.

Archer, R. P., Tirrell, C., & Elkins, D. E. (2001). Evaluation of an MMPI-A short form: Implications for adaptive testing. *Journal of Personality Assessment, 76,* 76–89.

Arizmendi, T. G., Beutler, L. E., Shanfield, S., Crago, M., & Hagaman, R. (1985). Client-therapist similarity and psychotherapy outcome: A microscopic approach. *Psychotherapy: Theory, Research, and Practice, 22,* 16–21.

Armbuster, G. L., Miller, A. S., & Exner, J. E. (1974). *Rorschach responses of parachute trainees at the beginning of training and shortly before their first jump.* Rorschach Workshops (Study No. 201, unpublished).

Armengol, C. G. (2001). The referral process. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 47–60). Lutz, FL: Psychological Assessment Resources.

Armengol, C. G., Moes, E. J., Penney, D. L., & Sapienza, M. M. (2001). Writing client-centered recommendations. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 141–160). Lutz, FL: Psychological Assessment Resources.

Army Individual Test Battery. (1944). *Manual of directions and scoring.* Washington, DC: War Department, Adjutant General's Office.

Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology, 20,* 112–119.

Arnold, M. B. (1949). A demonstration analysis of the TAT in a clinical setting. *Journal of Abnormal and Social Psychology, 44,* 97–111.

Arnold, M. B. (1962). *Story sequence analysis: A new method of measuring and predicting achievement.* New York: Columbia University Press.

Aronow, E., Reznikoff, M., & Moreland, K. L. (1995). The Rorschach: Projective technique or psychometric test? *Journal of Personality Assessment, 64,* 213–228.

Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology, 35,* 281–322.

Asaad, G. (2000). Somatization disorder. In M. Hersen & M. Biaggio (Eds.), *Effective brief therapies: A clinicians guide* (pp. 179–190). San Diego, CA: Academic Press.

Atkinson, J. W., & Feather, N. T. (Eds.). (1966). *A theory of achievement motivation.* New York: Wiley.

Atkinson, L. (1986). The comparative validities of the Rorschach and the MMPI: A meta-analysis. *Canadian Psychology, 27,* 238–347.

Atkinson, L., Quarington, B., Alp, I. E., & Cyr, J. J. (1986). Rorschach validity: An empirical approach to the literature. *Journal of Clinical Psychology, 42,* 360–362.

Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine, 15,* 433–442.

Axelrod, B. N. (2001). Administration duration for the Wechsler Adult Intelligence Scale-III and Wechsler Memory Scale-III. *Archives of Clinical Neuropsychology, 16,* 293–301.

Axelrod, B. N., Ryan, J. J., & Ward, L. C. (2001). Evaluation of seven-subtest short forms of the Wechsler Adult Intelligence Scale-III in a referred sample. *Archives of Clinical Neuropsychology, 16,* 1–8.

Axelrod, B. N., Ryan, J. J., & Woodward, J. L. (2001). Cross-validation of prediction equations for Wechsler Memory Scale-III. *Assessment, 8,* 367–372.

Axelrod, B. N., & Woodward, J. L. (2000). Parsiminous prediction of Wechsler Memory Scale-III memory indices. *Psychological Assessment, 12,* 431–435.

Baddeley, A. D., Wilson, B. A., & Watts, F. N. (1995). *Handbook of memory disorders.* New York: Wiley.

Baer, J. S., Holt, C. S., & Lichtenstein, E. (1986). Self-efficacy and smoking reexamined: Construct validity and clinical utility. *Journal of Consulting and Clinical Psychology, 54,* 846–852.

Baer, J. S., & Lichtenstein, E. (1988). Classification and prediction of smoking relapse episodes: An exploration of individual differences. *Journal of Consulting and Clinical Psychology, 56,* 104–110.

Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment, 72,* 308–320.

Baer, R. A., Wetter, M. W., & Berry, D. T. (1995). Effects of information about validity scales underreporting symptoms on the MMPI-2: An analogue investigation. *Assessment, 2,* 189–200.

Baer, R. A., Wetter, M. W., Nichols, D. S., Greene, R., & Berry, D. T. (1995). Sensitivity of MMPI-2 validity scales to underreporting of symptoms. *Psychological Assessment, 7,* 419–423.

Bagby, R. M., Buis, T., & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment, 7,* 84–92.

Bagby, R. M., Gillis, J. R., & Rogers, R. (1991). Effectiveness of the Millon Clinical Multiaxial Inventory Validity Index in the detection of random responding. *Psychological Assessment, 3,* 285–287.

Bagby, R. M., Gillis, J. R., Toner, B. B., & Goldberg, J. (1991). Detecting fake-good and fake-bad responding on the Millon Clinical Multiaxial Inventory-II. *Psychological Assessment, 3,* 496–498.

Bagby, R. M., Rogers, R., Nicholson, R. A., Buis, T., Seeman, M. V., & Rector, N. A. (1997). Effectiveness of the MMPI-2 validity indicators in the detection of

defensive responding in clinical and nonclinical samples. *Psychological Assessment, 9,* 406–413.

Bakker, C., Bakker-Rabdau, M., & Breit, S. (1978). The measurement of assertiveness and aggressiveness. *Journal of Personality Assessment, 42,* 277–284.

Ball, J. D., Archer, R. P., Gordon, R. A., & French, J. (1991). Rorschach Depression Indices with children and adolescents: Concurrent validity findings. *Journal of Personality Assessment, 57,* 465–476.

Ball, J. D., Archer, R. P., & Imhof, E. A. (1994). Time requirements of psychological testing: A survey of practitioners. *Journal of Personality Assessment, 63,* 239–249.

Bamgbose, O., Smith, G. T., Jesse, R. C., & Groth-Marnat, G. (1980). A survey of the current and future directions of professional psychology in acute general hospitals. *Clinical Psychologist, 33,* 24–25.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bannatyne, A. (1974). Diagnosis: A note on recategorization of the WISC scaled scores. *Journal of Learning Disabilities, 7,* 272–273.

Barber, T. X., & Silver, M. J. (1968). Fact, fiction and the experimenter bias effect. *Psychological Bulletin Monograph Supplement, 70,* 1–29.

Barlow, D. H. (1988). *Anxiety and its disorders: The nature and treatment of anxiety and panic.* New York: Guilford Press.

Barlow, D. H. (Ed.). (2001). *Clinical handbook of psychological disorders* (3rd ed.). New York: Guilford Press.

Barlow, D. H., Craske, M. G., Cerny, J. A., & Klosko, J. S. (1989). Behavioral treatment of panic disorder. *Behavior Therapy, 20,* 261–282.

Barnett, O. W., & Hamberger, L. K. (1992). The assessment of maritally violent men on the California Psychological Inventory. *Violence and Victims, 7,* 15–28.

Bar-On, R. (1998). *Bar-On Emotional Quotient Inventory (EQ-itm): Technical manual.* Toronto, Canada: Multi Health Systems.

Barona, A., Reynolds, C., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 26,* 74–75.

Barrios, B. A., & Hartman, D. P. (1988). Fears and anxieties in children. In E. J. Mash & L. G. Terdel (Eds.), *Behavioral assessment of childhood disorders* (2nd ed., pp. 196–262). New York: Guilford Press.

Barrios, B. A., & O'Dell, S. (1998). Fears and anxieties. In E. J. Mash & R. A. Barkley (Eds.), *Treatment of childhood disorders* (2nd ed., pp. 249–337). New York: Guilford Press.

Barthlow, D. L., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Incremental validity of the MMPI-2 content scales in an outpatient mental health setting. *Psychological Assessment, 11,* 39–47.

Bartlett, J. A., Schleifer, S. J., Johnson, R. L., & Keller, S. E. (1991). Depression in inner city adolescents attending an adolescent medicine clinic. *Journal of Adolescent Health, 12,* 316–318.

Basham, R. B. (1992). Clinical utility of the MMPI research scales in the assessment of adolescent acting out. *Psychological Assessment, 4,* 483–492.

Basso, M. B., Harrington, K., Matson, M., & Lowery, N. (2000). Sex differences on the WMS-III: Findings concerning verbal paired associates and faces. *Clinical Neuropsychologist, 14,* 231–235.

Baucom, D. H. (1985). Review of the California Psychological Inventory. In J. V. Mitchel (Ed.), *The ninth mental measurements yearbook.* Highland Park, NJ: Gryphon Press.

Baucom, D. H., Epstein, N., Rankin, L. A., & Burnett, C. K. (1996). Assessing relationship standards: The Inventory of Specific Relationship Standards. *Journal of Family Psychology, 10,* 72–88.

Baucom, D. H., Sayers, S. L., & Duhe, A. (1989). Marital attributions: Issues concerning attributional pattern and attributional style. *Journal of Personality and Social Psychology, 56,* 596–607.

Bauer, R. M. (2000). The flexible battery approach to neuropsychological assessment. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 419–448). Hillsdale, NJ: Erlbaum.

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects.* New York: Harper & Row.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York: Guilford Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II manual.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Garbin, M. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Beck, J. G. (1994). *Cognitive therapy: Basics and beyond.* New York: Guilford Press.

Beck, J. G., & Zebb, B. J. (1994). Behavioral assessment and treatment of panic disorder: Current status, future directions. *Behavior Therapy, 25,* 581–611.

Beck, S. J. (1937). Introduction to the Rorschach method: A manual of personality study. *American Orthopsychiatric Association Monograph, 1.*

Beck, S. J. (1945). *Rorschach's test: A variety of personality pictures* (Vol. 2). New York: Grune & Stratton.

Beck, S. J. (1951). The Rorschach Test: A multi-dimensional test of personality. In H. H. Anderson & G. Anderson (Eds.), *An introduction to projective techniques.* Englewood Cliffs, NJ: Prentice-Hall.

Beck, S. J. (1960). *The Rorschach experiment.* New York: Grune & Stratton.

Beck, S. J. (1961). *Rorschach's test: Basic processes* (Vol. 1). New York: Grune & Stratton.

Beck, S. J. (1968). Reality, Rorschach and perceptual theory. In A. I. Rabin (Ed.), *Projective techniques in personality assessment.* New York: Springer.

Beebe, D. W., Pfiffner, L. J., & McBurnett, K. (2000). Evaluation of the validity of the Wechsler Intelligence Scale for Children-3rd ed. Comprehension and Picture Arrangement subtests as measures of social intelligence. *Psychological Assessment, 12,* 97–101.

Beier, E. G. (1966). *The silent language of psychotherapy.* New York: Aldine.

Bellack, A. S., & Hersen, M. (Eds.). (1988). *Behavioral assessment: A practical handbook* (3rd ed.). New York: Pergamon Press.

Bellack, A. S., & Hersen, M. (Eds.). (1998). *Behavioral assessment: A practical handbook* (4th ed.). New York: Pergamon Press.

Bellack, A. S., Hersen, M., & Turner, S. M. (1979). Relationship of roleplaying and knowledge of appropriate behavior to assertion in the natural environment. *Journal of Consulting and Clinical Psychology, 47,* 670–678.

Bellak, L. (1954). *The Thematic Apperception Test and the Children's Apperception Test in clinical use.* New York: Grune & Stratton.

Bellak, L. (1975). *The TAT, CAT, and SAT in clinical use* (3rd ed.). New York: Grune & Stratton.

Bellak, L. (1986). *The TAT, CAT, and SAT in clinical use* (4th ed.). New York: Grune & Stratton.

Bellak, L. (1993). *The TAT, CAT, and SAT in clinical use* (5th ed.). New York: Grune & Stratton.

Bellak, L., & Abrams, D. M. (1997). *The TAT, CAT, and SAT in clinical use* (6th ed.). Boston: Allyn & Bacon.

Bellak, L., & Bellak, S. S. (1973). *Manual: Senior Apperception Test.* Larchmont, NY: CPS.

Bender, L. (1938). *A visual motor gestalt test and its clinical uses* (Research Monograms No. 3). New York: American Orthopsychiatric Association.

Ben-Porath, Y. S., & Butcher, J. N. (1989). The psychometric stability of rewritten MMPI items. *Journal of Personality Assessment, 53,* 645–653.

Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI-2 content scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment, 3,* 634–640.

Ben-Porath, Y. S., Hostetler, K., Butcher, J. N., & Graham, J. R. (1989). New subscales for the MMPI-2 Social Introversion (Si) Scale. *Psychological Assessment, 1,* 169–174.

Ben-Porath, Y. S., McCulley, E., & Almagor, M. (1993). Incremental validity of the MMPI-2 content scales in the assessment of personality and psychopathology by self-report. *Journal of Personality Assessment, 61,* 557–575.

Ben-Porath, Y. S., Shondrick, D. D., & Stafford, K. P. (1995). MMPI-2 and race in a forensic diagnostic sample. *Criminal Justice and Behavior, 22,* 19–32.

Benson, D. F. (1993). Aphasia. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (3rd ed., pp. 17–36). New York: Oxford University Press.

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia, 6,* 53–60.

Benton, A. L. (1974). *The Revised Visual Retention Test* (4th ed.). New York: Psychological Corporation.

Benton, A. L. (1979). Visuoperceptive, visuospatial, and visuoconstructive disorders. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 186–232). New York: Oxford University Press.

Benton, A. L., & Hamsher, K. (1983). *Multilingual Aphasia Examination.* Iowa City, IA: AJA Associates.

Benton, A. L., & Hamsher, K. (1989). *Multilingual Aphasia Examination* (2nd ed.). Iowa City, IA: AJA Associates.

Berg, F. W., Franzen, M. D., & Wedding, D. (1987). *Screening for brain impairment.* New York: Springer.

Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change.* New York: Wiley.

Berry, D. T., Baer, R. A., & Harris, M. J. (1991). Detection of malingering on the MMPI: A meta-analysis. *Journal of Personality Assessment, 11,* 585–598.

Beutler, L. E. (1979). Toward specific psychological therapies for specific conditions. *Journal of Consulting and Clinical Psychology, 47,* 882–897.

Beutler, L. E. (1983). *Eclectic psychotherapy: A systematic approach.* New York: Pergamon Press.

Beutler, L. E. (1989). Differential treatment selection: The role of diagnosis in psychotherapy. *Psychotherapy, 26,* 271–281.

Beutler, L. E. (1995). Integrating and communicating findings. In L. E. Beutler & M. R. Berren (Eds.), *Integrative assessment of adult personality* (pp. 25–64). New York: Guilford Press.

Beutler, L. E. (2000). David and Goliath: When empirical and clinical standards of practice meet. *American Psychologist, 55,* 997–1007.

Beutler, L. E., Albanese, A. L., Fisher, D., Karno, M., Sandowicz, M., Williams, O. B., et al. (1999). *Selecting and matching to patient variables.* Paper presented at the annual meeting of the Society for Psychotherapy Research, Braga, Portugal.

Beutler, L. E., & Berren, M. R. (Eds.). (1995). *Integrative assessment of adult personality.* New York: Guilford Press.

Beutler, L. E., & Clarkin, I. F. (1990). *Systematic treatment selection: Toward targeted therapeutic interventions.* New York: Brunner/Mazel.

Beutler, L. E., Clarkin, I. F., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patient.* New York: Oxford University Press.

Beutler, L. E., & Crago, M. (1986). Strategies and techniques of prescriptive psychotherapeutic intervention. In R. E. Hales & A. J. Francis (Eds.), *American Psychiatric Association Annual Review* (Vol. 6, pp. 378–397). Washington, DC: American Psychiatric Press.

Beutler, L. E., Crago, M., & Arizmendi, T. G. (1986). Therapist variables in psychotherapy process and outcome. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavioral change* (3rd ed., pp. 257–310). New York: Wiley.

Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., et al. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology, 59,* 1–8.

Beutler, L. E., & Groth-Marnat, G. (2003). *Integrative assessment of adult personality* (2nd ed.). New York: Guilford Press.

Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive psychotherapy.* New York: Oxford University Press.

Beutler, L. E., Harwood, T. M., Alimohamed, S., & Malik, M. (in press). Functional impairment and coping style. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient's needs.* New York: Oxford University Press.

Beutler, L. E., & Malik, M. (Eds.). (2002). *Rethinking the DSM: A psychological perspective.* Washington, DC: American Psychological Association.

Beutler, L. E., & Moleiro, C. (2001). Clinical versus reliable and significant change. *Clinical Psychology: Science and Practice, 8,* 441–445.

Beutler, L. E., Moleiro, C., Malik, M., & Harwood, T. M. (in press). A new twist on empirically supported treatments. *Revisita International de Psicologia Clinica y de la Salud.*

Beutler, L. E., Moleiro, C., & Talebi, H. (2002). Resistance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient's needs* (pp. 129–444). New York: Oxford University Press.

Beutler, L. E., Sandowicz, M., Fisher, D., & Albanese, A. L. (1996). Resistance in psychotherapy: What can be concluded from empirical research. *In Session: Psychotherapy in Practice, 2,* 77–86.

Beutler, L. E., & Williams, O. B. (1999). *Systematic Treatment Selection: A software package for treatment planning.* Ventura, CA: Center for Behavioral Technology.

Beutler, L. E., Williams, R. E., Wakefield, P. J., & Entwistle, S. R. (1995). Bridging scientist and practitioner perspectives in clinical psychology. *American Psychologist, 50,* 984–994.

Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The State-Trait Anxiety Inventory, Trait version: Structure and content re-examined. *Behaviour Research and Therapy, 36,* 777–788.

Bienvenu, O. J., Samuels, J. F., Riddle, M. A., Hoehn-Saric, R., Liang, K., Cullen, B. A. M., et al. (2000). The relationship of obsessive-compulsive disorder to possible spectrum disorders: Results from a family study. *Biological Psychiatry, 48,* 287–293.

Bigler, E. D., Rosa, L., Schultz, F., Hall, S., & Harris, J. (1989). Rey Auditory Verbal Learning and Rey-Osterrieth Complex Figure Design Test performance in Alzheimer's diseases and closed head injury. *Journal of Clinical Psychology, 45,* 277–280.

Billings, A. G., & Moos, R. H. (1986). Psychosocial processes of remission in unipolar depression: Comparing depressed patients with matched community controls. *Journal of Consulting and Clinical Psychology, 3,* 314–325.

Binder, R. L. (1999). Are the mentally ill dangerous? *Journal of the American Academy of Psychiatry and the Law, 27,* 189–201.

Birchler, G. R. (1989). Review of behavioral assessment: A practical handbook (3rd ed.). *Behavioral Assessment, 11,* 384–388.

Black, F. W., & Bernard, B. A. (1984). Constructional apraxia as a function of lesion locus and size in patients with focal brain damage. *Cortex, 20,* 111–120.

Blanchard, E. B., Schwarz, S. P., Neff, D. F., & Gerardi, M. A. (1988). Prediction of outcome from the self-regulatory treatment of irritable bowel syndrome. *Behavior Research and Therapy, 26,* 187–190.

Blatt, S. J., Zuroff, D. C., Quinlan, D. M., & Pilkonis, P. A. (1996). Interpersonal factors in brief treatment of depression: Further analysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64,* 162–171.

Blau, T. (1998). *The psychologist as expert witness* (2nd ed.). New York: Wiley.

Blum, G. S. (1950). *The Blacky pictures and manual.* New York: Psychological Corporation.

Blum, G. S. (1962). A guide for research use of the Blacky pictures. *Journal of Projective Techniques, 26,* 3–29.

Blum, G. S. (1968). Assessment of psychodynamic variables by the Blacky pictures. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 1, pp. 150–168). Palo Alto, CA: Science and Behavior Books.

Boccaccini, M. T., & Brodsky, S. L. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice, 30,* 252–259.

Bockian, N., Meagher, S., & Millon, T. (2000). Assessing personality with the Millon Behavioral Health Inventory, the Millon Behavioral Medicine Diagnostic, and the Millon Clinical Multiaxial Inventory. In R. J. Gatchel & J. N. Weisberg (Eds.), *Personality characteristics of patients with pain* (pp. 61–88). Washington, DC: American Psychological Association.

Boerger, A. R. (1975). *The utility of some alternative approaches to MMPI scale construction.* Unpublished doctoral dissertation, Kent State University, Kent, OH.

Boll, T. J. (1974). Behavioral correlates of cerebral damage in children age 9–14. In R. M. Reitan & L. A. Davison (Eds.), *Clinical neuropsychology: Current status and application.* Washington, DC: V. H. Winston & Sons.

Bolton, B. (1992). Review of the California Psychological Inventory, revised edition. In J. J. Framer & J. C. Conely (Eds.), *Eleventh mental measurements yearbook* (pp. 558–562). Lincoln, NE: Buros Institute of Mental Measurements.

Bonarius, H. (1984). Prediction or anticipation: Some implications of personal construct psychology for professional practice. *Jyvaskyla Studies in Education: Psychology and Social Research, 54,* 190–206.

Bornstein, R. A. (1983). Verbal I.Q.—Performance I.Q. discrepancies on the Wechsler Adult Intelligence Scale-Revised in patients with unilateral or bilateral cerebral dysfunction. *Journal of Consulting and Clinical Psychology, 51,* 779–789.

Bornstein, R. A. (1985). Normative data on selected neuropsychological measures from a nonclinical sample. *Journal of Clinical Psychology, 41,* 651–659.

Bornstein, R. A. (1986). Normative data on intermanual differences on three tests of motor performance. *Journal of Clinical and Experimental Neuropsychology, 8,* 12–20.

Bornstein, R. A., & Chelune, G. J. (1988). Factor analysis of the Wechsler Memory Scale-Revised. *Clinical Neuropsychologist, 2,* 107–115.

Bornstein, R. A., & Matarazzo, J. D. (1982). Wechsler VIQ versus PIQ differences in cerebral dysfunction: A literature review with emphasis on sex differences. *Journal of Clinical Neuropsychology, 4,* 319–334.

Bornstein, R. A., Suga, L., & Prifitera, A. (1987). Incidence of verbal I.Q.—Performance I.Q. discrepancies at various levels of education. *Journal of Clinical Psychology, 43,* 387–389.

Bornstein, R. F. (1998). Reconceptualizing personality disorder diagnosis in the *DSM-V:* The discriminant validity challenge. *Clinical Psychology: Science and Practice, 5,* 333–343.

Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment, 11,* 48–57.

Borus, J. F., Howes, M. J., Devins, N. P., & Rosenberg, R. (1988). Primary health care providers' recognition and diagnosis of mental disorders in their patients. *General Hospital Psychiatry, 10,* 317–321.

Botwinick, J., Storandt, M., Berg, F. W., & Boland, S. (1988). Senile dementia of the Alzheimer type: Subject attrition and testability in research. *Archives of Neurology, 45,* 493–496.

Bowden, S. C., Dodds, B., Whelan, G., Long, C., Dudgeon, P., Ritter, A., et al. (1997). Confirmatory factor analysis of the Wechsler Memory Scale-Revised in a sample of clients with alcohol dependence. *Journal of Clinical and Experimental Neuropsychology, 19,* 755–762.

Bowden, S. C., & Smith, L. C. (1994). What does the Austin Maze measure? *Australian Psychologist, 29,* 34–37.

Bowler, R. M., Thakler, C. D., & Becker, C. E. (1986). California Neuropsychological Screening Battery (CNC/BI & II). *Journal of Clinical Psychology, 42,* 946–955.

Bowman, M. L. (1996). Ecological validity of neuropsychological and other predictors following head injury. *Clinical Neuropsychologist, 10,* 382–396.

Bracken, B. A., & McCallum, R. S. (1998). *Examiner's manual: Universal Nonverbal Intelligence Test (UNIT).* Itasca, IL: Riverside.

Brar, H. S. (1970). Rorschach content responses of East Indian psychiatric patients. *Journal of Projective Techniques and Personality Assessment, 34,* 88–94.

Brawman-Mintzer, O., Lydiard, R. B., Emmanuel, N., Payeur, R., Johnson, M., Roberts, J., et al. (1993). Psychiatric comorbidity in patients with generalized anxiety disorder. *American Journal of Psychiatry, 150,* 1216–1218.

Brayfield, A. H. (Ed.). (1965). Testing and public policy. *American Psychologist, 20,* 857–1005.

Breen, M. J. (1982). Comparison of educationally handicapped student's scores on the Revised Developmental Test of Visual-Motor Integration and Bender-Gestalt. *Perceptual and Motor Skills, 54,* 1227–1230.

Breitholtz, E., Johansson, B., & Ost, L.-G. (1999). Cognitions in generalized anxiety disorder and panic disorder patients: A prospective approach. *Behavior Research and Therapy, 37,* 533–544.

Brewin, C. R. (1985). Depression and causal attributions: What is their relation? *Psychological Bulletin, 98,* 297–309.

Brewin, C. R. (1996). Theoretical foundations of cognitive-behavior therapy for anxiety and depression. *Annual Review of Psychology, 47,* 33–57.

Bricklin, H. (1984). *Bricklin Perceptual Scales.* Furlong, PA: Village Publishing.

Brink, T. L., Yesavage, J. A., Owen, L., Heersema, P. H., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontology, 1,* 37–43.

Brooks, B. L., & Hosie, T. W. (1984). Assumptions and interpretations of the SOMPA in estimating learning potential. *Counselor Education and Supervision, 23,* 290–299.

Broshek, D. K., & Barth, J. T. (2000). The Halstead-Reitan Neuropsychological Test Battery. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 457–532). New York: Wiley.

Broughton, R. (1984). A prototype strategy for construction of personality scales. *Journal of Personality and Social Psychology, 47,* 1334–1346.

Brown, C., Schulberg, H. C., & Madonia, M. J. (1995). Assessment of depression in primary care practice with the Beck Depression Inventory and the Hamilton Rating Scale for Depression. *Psychological Assessment, 7,* 59–65.

Brown, F. (1953). An exploratory study of dynamic factors in the content of the Rorschach protocol. *Journal of Projective Techniques, 17,* 251–279.

Brown, G. G., Spicer, K. B., & Robertson, W. M. (1989). Neuropsychological signs of lateralized arteriovenous malformations: Comparisons with ischemic stroke. *Clinical Neuropsychologist, 3,* 340–352.

Brown, L. (1990). Taking account of gender in the clinical assessment interview. *Professional Psychology, 21,* 12–17.

Brown, T. A., DiNardo, P. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV.* Albany, NY: Graywind.

Brown, T. A., O'Leary, T. A., & Barlow, D. H. (1993). Generalized anxiety disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (2nd ed., pp. 127–188). New York: Guilford Press.

Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *Journal of Applied Psychology, 87,* 185–191.

Bufka, L. F., Crawford, J. I., & Levitt, J. (2002). Brief screening instruments for managed care and primary care. In M. M. Antony & D. Barlow (Eds.), *Handbook of assessment, treatment planning, and outcome evaluation: Empirically supported strategies for psychological disorders* (pp. 49–62). New York: Guilford Press.

Buhler, C., & LeFever, D. (1947). A Rorschach study on the psychological characteristics of alcoholics. *Quarterly Journal of Studies on Alcoholism, 8,* 197–260.

Burns, G. (2001). *101 healing stories.* New York: Wiley.

Burt, D. B., Zembar, M. J., & Niederehe, G. (1995). Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin, 117,* 285–305.

Butcher, J. N. (1990). *The MMPI-2 in psychological treatment.* New York: Oxford University Press.

Butcher, J. N. (Ed.). (1996). *International adaptations of the MMPI-2.* Minneapolis: University of Minnesota Press.

Butcher, J. N. (1998). *Butcher Treatment Planning Inventory.* San Antonio, TX: Psychological Corporation.

Butcher, J. N. (1999). *A beginner's guide to the MMPI-2.* Washington, DC: American Psychological Association.

Butcher, J. N. (Ed.). (2000). *Basic sources of the MMPI-2.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for administration and scoring: MMPI-2.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 content scales.* Minneapolis: University of Minnesota Press.

Butcher, J. N., & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 25–50). Hillsdale, NJ: Erlbaum.

Butcher, J. N., & Hostetler, K. (1990). Abbreviating MMPI item administration: What can be learned from the MMPI for the MMPI-2? *Psychological Assessment, 2,* 12–21.

Butcher, J. N., Perry, J. N., & Atlis, M. M. (2000). Validity and utility of computer-based test interpretation. *Psychological Assessment, 12,* 6–18.

Butcher, J. N., & Pope, K. S. (1989). *MMPI-2: A practical guide to psychometric, clinical, and ethical issues.* Unpublished manuscript.

Butcher, J. N., & Williams, C. L. (1992). *Essentials of MMPI-2 and MMPI-A interpretation.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., et al. (1992). *MMPI-A (Minnesota Multiphasic Personality Inventory-Adolescent): Manual for administration, scoring, and interpretation.* Minneapolis: University of Minnesota Press.

Butler, M., Retzlaff, P., & Vanderploeg, R. (1991). Neuropsychological test usage. *Professional Psychology, 22,* 510–512.

Butters, N., Grant, I., Haxby, J., Judd, L. L., Martin, A., McClelland, J., et al. (1990). Assessment of AIDS-related cognitive changes: Recommendations of the NIMH

Workgroup on neuropsychological assessment approaches. *Journal of Clinical and Experimental Neuropsychology, 12,* 963–978.

Byrne, D., Golightly, C., & Sheffield, J. (1965). The Repression-Sensitization Scale as a measure of adjustment: Relationship with the CPI. *Journal of Consulting Psychology, 29,* 585–589.

Cacciola, J. S., Alterman, A. I., Rutherford, M. J., McKay, J. R., & May, D. J. (1999). Comparability of telephone and in-person structured clinical interview for *DSM-III-R* (SCID) diagnoses. *Assessment, 6,* 235–242.

Cacioppo, J. T., Glass, C. R., & Merluzzi, T. V. (1979). Self-statements and self-evaluations: A cognitive response analysis of heterosocial anxiety. *Cognitive Therapy and Research, 3,* 249–262.

Caldwell, A. B. (1988). *MMPI Supplemental Scale manual.* Los Angeles: Caldwell Report.

Caldwell, A. B. (2001). What do the MMPI scales fundamentally measure? Some hypotheses. *Journal of Personality Assessment, 76,* 1–17.

Callaghan, G. M. (2001). Demonstrating clinical effectiveness for individual practitioners and clinics. *Professional Psychology: Research and Practice, 32,* 289–297.

Calvin, J. (1975). *A replicated study of the concurrent validity of the Harris subscales for the MMPI.* Unpublished doctoral dissertation, Kent State University, Kent, OH.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31,* 141–154.

Campbell, J. (1998). Internal and external validity of seven Wechsler Intelligence Scale for Children (3rd ed., short forms in a sample of psychiatric inpatients). *Psychological Assessment, 10,* 431–434.

Campbell, K. A., Rohlunan, D. S., Storybach, D., Binder, L., Anger, W. K., Kovera, C. A., et al. (1999). Test-retest reliability of psychological and neurobehavioral tests set administered by computer. *Assessment, 6,* 21–32.

Canetti, L., Shalev, A. Y., & De-Nour, A. K. (1994). Israeli adolescent norms of the Brief Symptom Inventory (BSI). *Israel Journal of Psychiatry and Related Sciences, 31,* 13–18.

Canter, A. (1983). *The Canter Background Interference Procedure for the Bender Gestalt Test.* Los Angeles: Western Psychological Services.

Carkhuff, R. R. (1969). *Helping and human relations. I: Selection and training. II: Practice and research.* New York: Holt, Rinehart and Winston.

Carlson, J. G. (1982). Some concepts of perceived control and their relationship to bodily self-control. *Journal of Biofeedback and Self-Regulation, 7,* 341–375.

Carnes, G. D., & Bates, R. (1971). Rorschach anatomy response correlates in rehabilitation of failure subjects. *Journal of Personality Assessment, 35,* 527–537.

Cartwright, R. D. (1986). Affect and dream work from an information processing point of view. *Journal of Mind and Behavior, 7,* 411–428.

Caruso, J. C., & Cliff, N. (1999). The properties of equally and differentiated weighted WAIS-III factor scores. *Psychological Assessment, 11,* 198–206.

Caruso, K. R. (1988). *Manual for the Projective Storytelling Cards.* Sarasota, FL: Professional Resource Exchange.

Cass, W. A., & McReynolds, P. A. (1951). A contribution to Rorschach norms. *Journal of Consulting and Clinical Psychology, 15,* 178–183.

Cavalli, M., De Renzi, E., Faglioni, P., & Vitale, A. (1981). Impairment of right brain-damaged patients on a linguistic cognitive task. *Cortex, 17,* 545–556.

Cella, D. (1984). The modified WAIS-R: An extension and revision. *Journal of Clinical Psychology, 40,* 801–804.

Chambers, W. J., Puig-Antich, J., Hirsch, M., Paez, P., Ambrosini, P. J., Tabrizi, M. A., et al. (1985). The assessment of affective disorders in children and adolescents by semistructured interview: Test-retest reliability of the Schedule for Affective Disorders and Schizophrenia for School Age Children, Present Episode version. *Archives of General Psychiatry, 42,* 696–702.

Chase, T. N., Fedio, P., Foster, N. L., Brooks, R., DiChio, G., & Mansi, L. (1984). Wechsler Adult Intelligence Scale performance: Cortical localization by fluorodeoxyglucose F18-positive emission tomography. *Archives of Neurology, 41,* 1244–1247.

Chelune, G. J., & Bornstein, R. A. (1988). WMS-R patterns among patients with unilateral brain lesions [Special issue]. *Clinical Neuropsychologist, 2,* 121–132.

Chiles, J. A., Lambert, M. J., & Hatch, A. L. (1999). The impact of psychological interventions on medical cost offset: A meta-analytic review. *Clinical Psychology: Science and Practice, 6,* 204–220.

Chisholm, S. M., Crowther, J. H., & Ben-Porath, Y. S. (1997). Selected MMPI-2 Scales' ability to predict premature termination and outcome from psychotherapy. *Journal of Personality Assessment, 69,* 127–144.

Choca, J. P., Retzlaff, P., Strack, S., Mouton, A., & Van Denburg, E. (1996). Factorial elements in Millon's personality theory. *Journal of Personality Disorders, 10,* 377–383.

Choca, J. P., Shanley, L. A., Peterson, C. A., & Van Denburg, E. (1990). Racial bias and the MCMI. *Journal of Personality Assessment, 54,* 479–490.

Choca, J. P., & Van Denburg, E. (1997). *Interpretive guide to the Millon Clinical Multiaxial Inventory* (2nd ed.). Washington, DC: American Psychological Association.

Choca, J. P., Van Denburg, E., Bratu, M. E., & Meagher, S. (1995, March). *Personality changes of psychiatric patients with aging.* Paper presented at the midwinter meeting of the Society for Personality Assessment, Denver, CO.

Chojnacki, J. T., & Walsh, W. B. (1992). The consistency of scores and configural patterns between MMPI and MMPI-2. *Journal of Personality Assessment, 59,* 276–289.

Choquette, K. A. (1990). Assessing depression in alcoholics with the BDI, SCL-90-R, and DIS criteria. *Journal of Substance Abuse, 6,* 295–304.

Chusmir, L. H. (1985). Short-form scoring for McClelland's version of the TAT. *Perceptual and Motor Skills, 61,* 1047–1052.

Ciarrochi, J. V., Chan, A. Y. C., & Caputi, P. (2000). A critical evaluation of the emotional intelligence construct. *Personality and Individual Differences, 28*(3), 539–561.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290.

Ciminero, A. R., Calhoun, K. S., & Adams, H. E. (Eds.). (1977). *Handbook of behavioral assessment.* New York: Wiley.

Claassen, C. A., & Lovitt, R. (2001). Solving ethical problems in medical setting druing psychological assessment: A decisional model. *Journal of Personality Assessment, 77,* 214–230.

Clark, C. G., & Miller, H. L. (1971). Validation of Gilberstadt and Duker's 8–6 profile type on a black sample. *Psychological Reports, 29,* 259–264.

Clark, M. E. (1996). MMPI-2 Negative Treatment Indicators Content and Content Component Scales: Clinical correlates and outcome prediction for men with chronic pain. *Psychological Assessment, 8,* 32–38.

Clawson, A. (1962). *The Bender Visual Motor Gestalt for children: A manual.* Los Angeles: Western Psychological Services.

Coan, R. (1956). A factor analysis of Rorschach determinants. *Journal of Projective Techniques, 20,* 280–287.

Coche, E., & Sillitti, J. A. (1983). The Thematic Apperception Test as an outcome measure in psychotherapy. *Psychotherapy: Theory, Research, and Practice, 20,* 41–46.

Cojnacki, J. T., & Walsh, W. B. (1992). The consistency of scores and configural patterns between the MMPI and MMPI-2. *Journal of Personality Assessment, 59,* 278–289.

Colby, F. (1989). Usefulness of the K correction in MMPI profiles of patients and nonpatients. *Psychological Assessment, 1,* 142–145.

Colligan, R. C., & Offord, K. P. (1989). The aging MMPI: Contemporary norms for contemporary teenagers. *Mayo Clinic Proceedings, 64,* 3–27.

Collins, D. J. (1967). Psychological selection of drill sergeants: An exploratory attempt in a new program. *Military Medicine, 132,* 713–715.

Collins, J. M., & Bagozzi, R. P. (1999). Testing the equivalence of the socialization factor structure for criminals and noncriminals. *Journal of Personality Assessment, 72,* 68–73.

Collins, J. M., & Griffin, R. (1998). The psychology of underlying counterproductive job performance. In R. W. Griffin, A. O'Leary-Kelly, & J. M. Collins (Eds.), *Dysfunctional work behavior in organizations: Monographs in organizational behavior and industrial relations* (Vol. 23, Pt. B). Stamford, CT: JAL.

Cone, J. D. (1977). The relevance of reliability and validity for behavioral assessment. *Behavior Therapy, 8,* 411–426.

Cone, J. D. (1978). The behavioral assessment grid (BAG): A conceptual framework and taxonomy. *Behavior Therapy, 9,* 882–888.

Cone, J. D. (1998). Psychometric considerations: Concepts, contents, and methods. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (4th ed., pp. 22–46). Boston: Allyn & Bacon.

Conners, C. K. (1997). *Conners' Parent Rating Scale-Revised.* North Tonawanda, NY: Multi-Health Systems.

Conners, C. K. Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). Revision and restandardization of the Conners' Teacher Rating Scale (CTRS-R): Factor

structure reliability and criterion validity. *Journal of Abnormal Child Psychology, 26,* 279–291.

Cooney, N. L., Kadden, R. M., & Litt, M. D. (1990). A comparison of methods assessing sociopathy in male and female alcoholics. *Journal of Studies on Alcohol, 51,* 42–48.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87,* 442–449.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90,* 218–244.

Cooper, Z., & Fairburn, C. G. (1987). The eating disorder examination: A semistructured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders, 6,* 136–141.

Corcoran, K. (2000). *Measures for clinical practice: A sourcebook.* New York: Free Press.

Cordoni, B. K., O'Donnell, J. P., Ramaniah, N. V., Kurtz, J., & Rosenshein, K. (1981). Wechsler Adult Intelligence Scale patterns for learning disabled young adults. *Journal of Learning Disabilities, 14,* 404–407.

Corkin, S., Growdon, J. H., Sullivan, E. V., Nissen, M. J., & Huff, F. J. (1986). Assessing treatment effects: A neuropsychological battery. In L. W. Poon (Ed.), *Handbook for clinical memory assessment of older adults* (pp. 156–167). Washington, DC: American Psychological Association.

Cormier, W. H., & Cormier, L. S. (1998). *Interviewing strategies for helpers: Fundamental skills and cognitive behavioral interventions* (4th ed.). Monterey, CA: Brooks/Cole.

Coryell, W. H., Akiskal, H. S., Leon, A. C., Winokur, G., Masur, J. D., Mueller, T. I., et al. (1994). The time course of nonchronic major depressive disorder. *Archives of General Psychiatry, 51,* 405–410.

Costa, P. T., Zonderman, A. B., Williams, R. B., & McRae, R. R. (1985). Content and comprehensiveness of the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology, 48,* 925–933.

Costantino, G., & Malgady, R. G. (1999). The Tell-Me-A-Story Test: A multicultural offspring of the Thematic Apperception Test. In L. Gieser & M. I. Stein (Eds.), *Evocative images: The Thematic Apperception Test and the art of projection* (pp. 177–190). Washington, DC: American Psychological Association.

Costantino, G., Malgady, R. G., Casullo, M. M., & Castillo, A. (1991). Cross-cultural standardization of the TEMAS in three Hispanic subcultures. *Journal of Behavioral Sciences, 13,* 48–62.

Costantino, G., Malgady, R. G., Colon-Malgady, G., & Bailey, J. (1992). Clinical utility of the TEMAS with nonminority children. *Journal of Personality Assessment, 59,* 433–438.

Costantino, G., Malgady, R. G., & Rogler, L. H. (1988). *Technical manual: TEMAS Thematic Apperception Test.* Los Angeles: Western Psychological Services.

Costantino, G., Malgady, R. G., Rogler, L. H., & Tsui, E. C. (1988). Discriminant analysis of clinical outpatients and public school children by TEMAS: A Thematic Apperception Test for Hispanics and Blacks. *Journal of Personality Assessment, 52,* 670–678.

Costello, C. G. (1958). The Rorschach records of suicidal patients. *Journal of Projective Techniques, 22,* 272–275.

Costello, E. J., Benjamin, R., Angold, A., & Silver, D. (1991). Mood variability in adolescence: A study of depressed and nondepressed comorbid patients. *Journal of Affective Disorders, 23,* 199–212.

Costello, E. J., Edelbrock, C. S., & Costello, A. J. (1985). Validity of the NIMH Diagnostic Interview Schedule for Children: A comparison between psychiatric and pediatric referrals. *Journal of Abnormal and Child Psychology, 13,* 579–595.

Costello, E. J., Edelbrock, C. S., Duncan, M. K., & Kalas, R. (1984). *Testing of the NIMH Diagnostic Interview Schedule for Children (DISC) in a clinical population. Final report to the Center for Epidemiological Studies, National Institute for Mental Health.* Pittsburgh, PA: University of Pittsburgh.

Costello, R. M., & Wicott, K. A. (1984). Impression management and testing for locus of control in an alcoholic sample. *International Journal of the Addictions, 19,* 45–56.

Craig, R. J. (1995). Clinical diagnoses and MCMI codetypes. *Journal of Clinical Psychology, 51,* 352–360.

Craig, R. J. (1999). Overview and status of the Millon Clinical Multiaxial Inventory. *Journal of Personality Assessment, 72,* 390–406.

Craig, R. J. (2001). Adjectival descriptions of personality disorders: A convergent validity study of the MCMI-III. *Journal of Personality Assessment, 77,* 259–271.

Craig, R. J., & Bivens, A. (1998). Factor structure of the MCMI-III. *Journal of Personality Assessment, 70,* 190–196.

Cramer, P. (1996). *Storytelling, narrative, and the Thematic Apperception Test.* New York: Guilford Press.

Cramer, P. (1998). Threat to gender representation: Identity and identification. *Journal of Personality Assessment, 66,* 335–357.

Cramer, P. (1999). Future directions for the Thematic Apperception Test. *Journal of Personality Assessment, 72,* 74–92.

Crary, W. G., & Johnson, C. W. (1981). Mental status examination. In C. W. Johnson, J. R. Snibbe, & L. A. Evans (Eds.), *Basic psychopathology: A programmed text* (2nd ed., pp. 56–57). Lancaster, PA: MTP Press.

Craske, M. G., & Barlow, D. H. (1993). Panic and agoraphobia. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (pp. 1–47). New York: Guilford Press.

Crawford, J. R., Millar, J., & Milne, A. B. (2001). Estimating premorbid IQ from demographic variables: A comparison of a regression equation vs. clinical judgment. *British Journal of Clinical Psychology, 40,* 97–105.

Crawford, J. R., Stewart, L. E., & Moore, J. W. (1989). Demonstration of savings on the AVLT and development of a parallel form. *Journal of Clinical and Experimental Neuropsychology, 11,* 975–981.

Crockett, D., Tallman, K., Hurwitz, T., & Kozak, J. (1988). Neuropsychological performance in psychiatric patients with or without documented brain dysfunction. *International Journal of Neuroscience, 41,* 71–79.

Cronbach, L. J. (1978). Black Intelligence Test of Cultural Homogeneity: A review. In O. K. Buros (Ed.), *The eighth mental measurements yearbook* (Vol. 1, p. 250). Highland Park, NJ: Gryphon Press.

Crumpton, E. (1956). The influence of color on the Rorschach Test. *Journal of Projective Techniques, 20,* 150–158.

Cummings, N. A. (1991). The somatizing patient. In C. A. Austad & W. A. Berman (Eds.), *Psychotherapy in managed care: The optimal use of time and resources* (pp. 234–237). Washington, DC: American Psychological Association.

Cummings, N. A. (1999). Medical cost offset, meta-analysis, and implications for future research and practice. *Clinical Psychology: Science and Practice, 6,* 221–224.

Cunningham, T. R., & Thorp, R. G. (1981). The influence of settings on accuracy and reliability of behavioral observation. *Behavioral Assessment, 3,* 67–78.

Cyr, J. J., McKenna-Foley, J. M., & Peacock, E. (1985). Factor structure of the SCL-90-R: Is there one? *Journal of Personality Assessment, 49,* 571–578.

Dahlstrom, W. G. (1969). Recurrent issues in the development of the MMPI. In J. N. Butcher (Ed.), *MMPI: Research developments and clinical applications.* New York: McGraw-Hill.

Dahlstrom, W. G., & Welsh, G. S. (1960). *An MMPI handbook: A guide to use in clinical practice and research.* Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook. Vol. 1: Clinical Interpretation.* Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1975). *An MMPI handbook. Vol. 2: Research developments and applications.* Minneapolis: University of Minnesota Press.

Daldrup, F. J., Beutler, L. E., Engle, D., & Greenberg, L. S. (1988). *Focused expressive psychotherapy: Freeing the overcontrolled patient.* New York: Pergamon Press.

Dana, R. H. (1955). Clinical diagnosis and objective TAT scoring. *Journal of Abnormal and Social Psychology, 50,* 19–24.

Dana, R. H. (1965). Review of the Rorschach. In O. K. Buros (Ed.), *Sixth mental measurements yearbook* (pp. 492–495). Highland Park, NJ: Gryphon Press.

Dana, R. H. (1968). Six constructs to define Rorschach M. *Journal of Projective Techniques and Personality Assessment, 32,* 138–145.

Dana, R. H. (1999). Cross-cultural-multimethod use of the Thematic Apperception Test. In L. Gieser & M. I. Stein (Eds.), *Evocative images: The Thematic Apperception Test and the art of projection* (pp. 177–190). Washington, DC: American Psychological Association.

Dana, R. H. (2000). *Handbook of cross-cultural and multi-cultural personality assessment.* Mahwah, NJ: Erlbaum.

Dana, R. H., & Cocking, R. R. (1968). Cue parameters, cue probabilities, and clinical judgment. *Journal of Clinical Psychology, 24,* 475–480.

Dana, R. H., Field, K., & Bolton, B. (1983). Variations of the Bender-Gestalt Test: Implications for training and practice. *Journal of Personality Assessment, 47,* 76–84.

Daniel, M. H. (1997). Intelligence testing: Status and trends. *American Psychologist, 52,* 1038–1045.

Das, J. P., & Naglieri, J. A. (1994). *Das-Naglieri Cognitive Assessment System: Standardization Test Battery.* Chicago: Riverside Press.

Datel, W. E., Hall, F. D., & Rufe, C. P. (1965). Measurement of achievement motivation in army security agency foreign language candidates. *Educational and Psychological Measurement, 25,* 539–545.

Davis, G. L., Hoffman, R. G., & Nelson, K. S. (1990). Differences between Native Americans and Whites on the California Psychological Inventory. *Psychological Assessment, 2,* 238–242.

Davis, R. G. (1999). Millon: Essentials of his science, theory, classification, assessment, and therapy. *Journal of Personality Assessment, 72,* 330–352.

Davis, R. G., Wenger, A., & Guzman, A. (1997). Validation of the MCMI-III. In T. Millon (Ed.), *The Millon inventories: Clinical and personality assessment* (pp. 327–362). New York: Guilford Press.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95–106.

Dean, R. S. (1982). Neuropsychological assessment. In T. Kratochwill (Ed.), *Advances in school psychology* (Vol. 2). Hillsdale, NJ: Erlbaum.

Deardorff, W. W. (2000). The MMPI-2 and chronic pain. In R. J. Gatchel & J. N. Weisberg (Eds.), *Personality characteristics of patients with pain* (pp. 109–125). Washington, DC: American Psychological Association.

Debord, J. B. (1989). Paradoxical interventions: A review of the recent literature. *Journal of Counseling and Development, 67,* 394–398.

DeFrancesco, J. J., & Taylor, J. (1993). A validation on the revised Socialization Scale of the California Psychological Inventory. *Journal of Psychopathology and Behavioral Assessment, 15,* 53–56.

DeGood, D. E., Crawford, A. L., & Jongsma, A. E. (1999). *The behavioral medicine treatment planner.* New York: Wiley.

Del Greco, L., Breitbach, L., & McCarthy, R. H. (1981). The Rathus Assertiveness Schedule modified for early adolescents. *Journal of Behavioral Assessment, 3,* 321–328.

Delis, D. C., Kaplan, E., & Kramer, J. H. (1999). *Delis-Kaplan Executive Function System.* San Antonio, TX: Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult version.* San Antonio, TX: Psychological Corporation.

Deniston, W. M., & Ramanaiah, N. V. (1993). California Psychological Inventory and the five-factor model of personality. *Psychological Reports, 73,* 491–496.

Denman, S. (1984). *Denman Neuropsychology Memory Scale.* Charleston, SC: Author.

Derogatis, L. R. (1992). *BSI: Administration, scoring, and procedures manual-II* (2nd ed.). Baltimore: Clinical Psychometric Research.

Derogatis, L. R. (1993). *Brief Symptom Inventory (BSI) administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems.

Derogatis, L. R. (1994). *SCL-90-R: Administration, scoring, and procedures manual.* Minneapolis, MN: National Computer Systems.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Scientist, 19,* 1–15.

Derogatis, L. R., & Savitz, K. L. (1999). The SCL-90-R, Brief Symptom Inventory, and matching clinical rating scales. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 679–724). Mahwah, NJ: Erlbaum.

desRosiers, G. (1992). Primary or depressive dementia: Clinical features. *International Journal of Geriatric Psychiatry, 7,* 629–638.

desRosiers, G., & Kavanagh, D. (1987). Cognitive assessment in closed head injury: Stability, validity, and parallel forms for two neuropsychological measures of recovery. *International Journal of Clinical Neuropsychology, 9,* 162–173.

Dicken, C. F. (1960). Simulated patterns on the California Psychological Inventory. *Journal of Counseling Psychology, 7,* 24–31.

DiClemente, C. (1986). Self-efficacy and the addictive behaviors. *Journal of Social and Clinical Psychology, 4,* 302–315.

Dies, R. R. (1972). Personal gullibility or pseudodiagnosis: A further test of the "fallacy of personal validation." *Journal of Clinical Psychology, 28,* 47–50.

Diller, L., Ben-Yishay, Y., Gertsman, L. J., Goodkin, R., Gordon, W., & Weinberg, J. (1976). *Studies in cognition and rehabilitation in hemiplegia* (Rehabilitation Monograph, No. 50). New York: New York University Medical Center, Institute of Rehabilitation Medicine.

Dixon, W. E., & Anderson, T. (1995). Establishing covariance continuity between the WISC-R and the WISC-III. *Psychological Assessment, 7,* 115–117.

Dobbins, C., & Russell, E. W. (1990). Left temporal lobe brain damage pattern on the Wechsler Adult Intelligence Scale. *Journal of Clinical Psychology, 46,* 863–868.

Dodrill, C. B., & Troupin, A. S. (1975). Effects of repeated administration of a comprehensive neuropsychological battery among chronic epileptics. *Journal of Nervous and Mental Diseases, 161,* 185–190.

Dolman, R., Roy, E. A., Dimeck, P. T., & Hall, C. R. (2000). Age, gesture span, and dissociation among component subsystems of working memory. *Brain, Cognition, and Rehabilitation, 43,*164–168.

Dorr, D. (1999). Approaching psychotherapy of the personality disorders from the Millon perspective. *Journal of Personality Assessment, 72,* 407–425.

Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology, 71,* 9–15.

Dove, A. (1968). Taking the Chitling Test. *Newsweek, 72,* 51–52.

Dowd, E. T., Milne, C. R., & Wise, S. L. (1991). The Therapeutic Reactance Scale: A measure of psychological reactance. *Journal of Counseling and Development, 69,* 541–545.

Dowd, E. T., & Wallbrown, F. (1993). Motivational components of client reactance. *Journal of Counseling and Development, 71,* 533–538.

Dozois, D. J., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10,* 83–89.

Draguns, J. G., Haley, E. M., & Phillips, L. (1967). Studies of the Rorschach content: A review of the research literature. Part 1: Traditional content categories. *Journal of Projective Techniques and Personality Assessment, 31,* 3–32.

Drake, L. E. (1946). A social I. E. scale for the MMPI. *Journal of Applied Psychology, 30,* 51–54.

Dritschel, B. H., Williams, K., & Cooper, P. J. (1991). Cognitive distortions among women experiencing bulimic episodes. *International Journal of Eating Disorders, 10,* 547–555.

DuAlba, L., & Scott, R. L. (1993). Somatization and malingering for workers' compensation applicants: A cross-cultural MMPI study. *Journal of Clinical Psychology, 49,* 913–917.

Dudek, S. Z. (1968). M and active energy system correlating Rorschach M with ease of creative expression. *Journal of Projective Techniques and Personality Assessment, 32,* 453–461.

Durand, V. M. (1990). *Severe behavior problems: A functional communication training approach* (pp. 80–82). New York: Guilford Press.

Duthie, B., & Vincent, K. R. (1986). Diagnostic hit rates of high point codes for the Diagnostic Inventory of Personality and Symptoms using random assignment, base rates, and probability scales. *Journal of Clinical Psychology, 42,* 612–614.

Dye, O. A. (1979). Effects of practice on Trail Making performance. *Perceptual and Motor Skills, 48,* 296.

Earls, R., Reich, W., Jung, K. G., & Cloninger, C. R. (1988). Psychopathology in children of alcoholic and antisocial parents. *Alcoholism: Clinical and Experimental Research, 12,* 481–487.

Eaton, W. W., Neufeld, K., Chen, L., & Cai, G. (2000). A comparison of self report and clinical diagnostic interviews for depression. *Archives of General Psychiatry, 57,* 217–222.

Edell, W. S. (1987). Role of structure in disordered thinking in borderline and schizophrenic disorders. *Journal of Personality Assessment, 51,* 23–41.

Edwards, A. L. (1957). *The social desirability variables in personality assessment and research*. New York: Dryden Press.

Edwards, A. L. (1964). Social desirability and performance on the MMPI. *Psychometrika, 29,* 295–308.

Edwards, D. W., Morrison, T. L., & Weissman, H. N. (1993). The MMPI and MMPI-2 in an outpatient sample: Comparisons of code types, validity scales, and clinical scales. *Journal of Personality Assessment, 61,* 1–18.

Egeland, B. R. (1969). Examiner expectancy: Effects on the scoring of the WISC. *Psychology in the Schools, 6,* 313–315.

Eichler, R. M. (1951). Experimental stress and alleged Rorschach indices of anxiety. *Journal of Abnormal and Social Psychology, 46,* 169–177.

Eidelson, R. J., & Epstein, N. (1982). Cognitive and relationship adjustment: Development of a measure of dysfunctional relationship beliefs. *Journal of Consulting and Clinical Psychology, 50,* 715–720.

Eisman, E. J., Dies, R. R., Finn, S. E., Eyde, L. D., Kay, G. G., Kubiszyn, T. W., et al. (2000). Problems and limitations in using psychological assessment in the contemporary health care delivery system. *Professional Psychology: Research and Practice, 31,* 131–140.

Elashoff, J., & Snow, R. E. (Eds.). (1971). *Pygmalion revisited.* Worthington, OH: C. A. Jones.

Elion, V. H., & Megargee, E. I. (1975). Validity of the MMPI Pd Scale among black males. *Journal of Consulting and Clinical Psychology, 43,* 166–172.

Elkin, I., Shea, T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., et al. (1989). National Institute of Mental Health treatment of depression collaborative research program. *Archives of General Psychiatry, 46,* 971–982.

Elliot, L. L. (1960). *WAF performance on the California Psychological Inventory* (Wright Air Development Division Technical Note 60–218). San Antonio, TX: Air Research and Development Command, Lackland AFB.

Ellis, M. V., Robbins, E. S., Schult, D., Ladny, N., & Baker, J. (1990). Anchoring errors in clinical judgments: Type I error, adjustment or mitigation. *Journal of Counseling Psychology, 37,* 343–351.

Elstein, A. A. (1965). Behavioral correlates of the Rorschach shading determinant. *Journal of Consulting Psychology, 29,* 231–236.

Elwood, R. W. (1991). The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application. *Neuropsychology Review, 2,* 179–201.

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry, 35,* 837–844.

Entwisle, D. R. (1972). To dispel fantasies about fantasy-based measures of achievement motivation. *Psychological Bulletin, 77,* 377–391.

Eppinger, M. G., Craig, P. L., Adams, R. L., & Parsons, O. A. (1987). The WAIS-R Index for estimating premorbid intelligence: Cross-validation and clinical utility. *Journal of Consulting and Clinical Psychology, 55,* 86–90.

Epstein, J., & Klinkenberg, W. D. (2001). From Eliza to Internet: A brief history of computerized assessment. *Computers in Human Behavior, 17,* 295–314.

Equal Employment Opportunity Commission. (1970). Guidelines on employee selection procedures. *Federal Register, 35,* 12333–12336.

Erginel, A. (1972). On the test-retest reliability of the Rorschach. *Journal of Personality Assessment, 36,* 203–212.

Erickson, R. C., Caslyn, D. A., & Scheupbach, C. S. (1978). Abbreviating the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology, 42,* 946–955.

Eron, L. D. (1950). A normative study of the Thematic Apperception Test. *Psychological Monographs, 64,* 315.

Eslinger, P. J., Damasio, A. R., & Benton, A. L. (1984). *The Iowa Screening Battery for Mental Decline.* Iowa City: University of Iowa.

Exner, J. E. (1969). *The Rorschach systems.* New York: Grune & Stratton.

Exner, J. E. (1974). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* New York: Wiley.

Exner, J. E. (1978). *The Rorschach: A comprehensive system. Volume 2: Current research and advanced interpretations.* New York: Wiley.

Exner, J. E. (1979). *The effects of voluntary restraint on Rorschach retests.* Rorschach Workshops (Study No. 258, unpublished).

Exner, J. E. (1983). Rorschach assessment. In I. B. Weiner (Ed.), *Clinical methods in clinical psychology* (2nd ed.). New York: Wiley.

Exner, J. E. (1984). *A computer program to assist in Rorschach interpretation* (Rev. ed.). Bayville, NY: Rorschach Workshops.

Exner, J. E. (1986). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (2nd ed.). New York: Wiley.

Exner, J. E. (1988). Problems with brief Rorschach protocols. *Journal of Personality Assessment, 52,* 640–647.

Exner, J. E. (1991). *The Rorschach: A comprehensive system. Volume 2: Current research and advanced interpretation* (2nd ed.). New York: Wiley.

Exner, J. E. (1993). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (3rd ed.). New York: Wiley.

Exner, J. E. (1995). *The Rorschach: A comprehensive system. Volume 3: Assessing children and adolescents.* New York: Wiley.

Exner, J. E. (1997). The future of Rorschach in personality assessment. *Journal of Personality Assessment, 68,* 37–46.

Exner, J. E. (2000). *A primer for Rorschach interpretation.* Asheville, NC: Rorschach Workshops.

Exner, J. E. (2001). *A Rorschach workbook for the comprehensive system* (5th ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E. (2003). *The Rorschach: A comprehensive system. Volume 1: Basic foundations* (4th ed.). New York: Wiley.

Exner, J. E., Armbuster, G. L., & Viglione, D. (1978). The temporal stability of some Rorschach features. *Journal of Personality Assessment, 42,* 474–482.

Exner, J. E., Armbuster, G. L., Walker, E. J., & Cooper, W. H. (1975). *Anticipation of elective surgery as manifest in Rorschach records.* Rorschach Workshops (Study No. 213, unpublished).

Exner, J. E., & Bryant, E. L. (1974). *Rorschach responses of subjects recently divorced or separated.* Rorschach Workshops (Study No. 206, unpublished).

Exner, J. E., & Exner, D. E. (1972). How clinicians use the Rorschach. *Journal of Personality Assessment, 36,* 403–408.

Exner, J. E., & Hillman, L. (1984). *A comparison of content distributions for the records of 76 paranoid schizophrenics and 76 nonparanoid schizophrenics.* Rorschach Workshops (Study No. 293, unpublished).

Exner, J. E., & Kazaoka, K. (1978). *Dependency gestures of 16 assertiveness trainees as related to Rorschach movement responses.* Rorschach Workshops (Study No. 261, unpublished).

Exner, J. E., Leura, A. V., & George, L. M. (1976). *A replication of the Masling study using four groups of new examiners with two seating arrangements and ride evaluation.* Rorschach Workshops (Study No. 256, unpublished).

Exner, J. E., & Murillo, L. G. (1973). Effectiveness of regressive ECT with process schizophrenics. *Diseases of the Nervous System, 34,* 44–48.

Exner, J. E., & Murillo, L. G. (1977). A long-term follow up of schizophrenics treated with regressive ECT. *Diseases of the Nervous System, 38,* 162–168.

Exner, J. E., Murillo, L. G., & Sternklar, S. (1979). *Anatomy and X-ray responses among patients with body delusions or body problems.* Rorschach Workshops (Study No. 257, unpublished).

Exner, J. E., & Sanglade, A. A. (1992). Rorschach changes following brief and short term therapy. *Journal of Personality Assessment, 59,* 59–71.

Exner, J. E., Thomas, E. A., Cohen, J. B., Ridgeway, E. M., & Cooper, W. H. (1981). *Stress indices in the Rorschachs of patients recovering from myocardial infarctions.* Rorschach Workshops (Study No. 286, unpublished).

Exner, J. E., Thomas, E. A., & Martin, L. S. (1980). *Alterations in G. S. R. and cardiac and respiratory rates in introversives and extratensives during problem solving.* Rorschach Workshops (Study No. 272, unpublished).

Exner, J. E., Thomas, E. A., & Mason, B. (1985). Children's Rorschachs: Descriptions and prediction. *Journal of Personality Assessment, 49,* 13–20.

Exner, J. E., & Weiner, I. B. (1982). *The Rorschach: A comprehensive system. Volume 3: Assessment of children and adolescents.* New York: Wiley.

Exner, J. E., & Weiner, I. B. (1995). *The Rorschach: A comprehensive system. Volume 3: Assessment of children and adolescents* (2nd ed.). New York: Wiley.

Exner, J. E., & Wylie, J. R. (1975). *Attempts at simulation of schizophrenia-like protocols by psychology graduate students* (Workshops Study No. 211, unpublished). Bayville, NY: Rorschach Workshops.

Exner, J. E., & Wylie, J. R. (1977). Some Rorschach data concerning suicide. *Journal of Personality Assessment, 41,* 339–348.

Exner, J. E., Wylie, J. R., Leura, A. V., & Parrill, T. (1977). Some psychological characteristics of prostitutes. *Journal of Personality Assessment, 41,* 474–485.

Exner, J. E., Zalis, T., & Schumacher, J. (1976). *Rorschach protocols of chronic amphetamine users.* Rorschach Workshops (Study No. 233, unpublished).

Eysenck, H. J. (1985). Review of the California Psychological Inventory. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 252–253). Highland Press, NJ: Gryphon Press.

Eysenck, H. J. (1994). The outcome problem in psychotherapy: What have we learned? *Behavior Research and Therapy, 32,* 477–495.

Fals-Stewart, W. (1995). The effect of defensive responding by substance-abusing patients on the Millon Clinical Multiaxial Inventory. *Journal of Personality Assessment, 64,* 540–541.

Farr, S. P., & Martin, P. W. (1988). Neuropsychological dysfunction. In R. L. Greene (Ed.), *The MMPI: Use with special populations* (pp. 214–245). San Diego, CA: Grune & Stratton.

Faust, D. (1991). Forensic neuropsychology: The art of practicing a science that does not yet exist. *Neuropsychology Review, 2,* 205–231.

Faust, D., & Ziskin, J. (1989). Computer-assisted psychological evaluation as legal evidence: Someday my prints will come. *Computers in Human Behavior, 5,* 23–36.

Faust, D., Ziskin, J., & Hiers, J. B. (1991). *Brain damage claims: Coping with neuropsychological evidence* (Vols. 1–2). Los Angeles: Law and Psychology Press.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26,* 57–63.

Feingold, A. (1983). The validity of the information and vocabulary subtests of the WAIS for predicting college achievement. *Educational and Psychological Measurement, 43,* 1127–1131.

Feldman, S. E., & Sullivan, D. S. (1971). Factors mediating the efforts of enhanced rapport on children's performance. *Journal of Consulting and Clinical Psychology, 36,* 302.

Fernandez-Ballesteros, R., & Staats, A. W. (1992). Paradigmatic behavioral assessment, treatments, and evaluation: Answering the crisis in behavioral assessment. *Advances in Behavior Research and Therapy, 14,* 1–27.

Filskov, S. B. (1978). *The prediction of impairment from figure copying.* Paper presented at the Southeastern Psychological Association convention, Atlanta, GA.

Fincham, F. D., & Bradbury, T. N. (1992). Assessing attributions in marriage: The Relationship Attribution Measure. *Journal of Personality and Social Psychology, 62,* 457–468.

Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment, 11,* 58–66.

Finlayson, M. A. J., & Bird, D. R. (1991). Psychopathology and neuropsychological deficit. In H. O. Doerr & A. S. Carlin (Eds.), *Forensic neuropsychology: Legal and scientific basis* (pp. 123–140). New York: Guilford Press.

Finlayson, M. A. J., & Reitan, R. M. (1980). Effect of lateralized lesions on ipsilateral and contralateral motor functioning. *Journal of Clinical Neuropsychology, 2,* 237–243.

Finn, S. E. (1982). Base rates, utilities, and the *DSM-III*: Shortcomings of fixed rule systems of psychodiagnostics. *Journal of Abnormal Psychology, 48,* 294–302.

Finn, S. E. (1996). *A manual for using the MMPI-2 as a therapeutic intervention.* Minneapolis: University of Minnesota Press.

Finn, S. E., Moes, E. J., & Kaplan, E. (2001). The consumer's point of view. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 13–46). Lutz, FL: Psychological Assessment Resources.

Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment, 4,* 278–287.

Finn, S. E., & Tonsager, M. E. (1997). Information gathering and therapeutic models of assessment: Complementary paradigms. *Psychological Assessment, 9,* 374–385.

First, M. B., Frances, A., Widiger, T. A., Pincus, H. A., & Davis, W. W. (1992). *DSM-IV* and behavioral assessment. *Behavioral Assessment, 14,* 297–306.

First, M. B., Gibbon, M., Williams, J. B., & Spitzer, R. L. (1996). *Users manual for the Auto SCID-II (for DSM-IV).* North Tonawanda, NY: Multi-Health Systems.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). The Structured Clinical Interview for *DSM-III-R* Personality disorders (SCID-II): Description. *Journal of Personality Disorders, 9,* 83–91.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for Axis I DSM-IV: Disorders Research version-Patient edition (SCID-I/P, Version 2.0).* New York: New York State Psychiatric Institute, Biometrics Research Department.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)-Clinician version.* Washington, DC: American Psychiatric Press.

Fisher, D. C., Beutler, L. E., & Williams, O. B. (1999). Making assessment relevant to treatment planning: The STS Clinician Rating Form. *Journal of Clinical Psychology, 55,* 825–842.

Fisher, D. C., Ledbetter, M. F., Cohen, N. J., Marmor, D., & Tulsky, D. S. (2000). WAIS-III and WMS-III profiles of mildly to severely brain-injured patients. *Applied Neuropsychology, 7,* 126–132.

Fisher, S. (1962). Relationship of Rorschach human percepts to projective descriptions with self reference. *Journal of Projective Techniques, 26,* 231–233.

Flaherty, M. R., & Reutzel, G. (1965). Personality traits of high and low achievers in college. *Journal of Educational Research, 58,* 409–411.

Flanagan, D. P., Genshaft, J. L., & Harrison, P. L. (Eds.). (1997). *Contemporary intellectual assessment.* New York: Guilford Press.

Flanagan, R., & Guiseppe, R. D. (1999). Critical review of the TEMAS: A step within the development of thematic apperception instruments. *Psychology in the Schools, 36,* 21–30.

Flaugher, R. L., & Schrader, W. B. (1978). *Eliminating differentially difficult items as an approach to test bias (RB-78–4).* Princeton, NJ: Educational Testing Service.

Flynn, P. M., McCann, J. T., & Fairbank, J. A. (1995). Issues in the assessment of personality disorder and substance abuse using the Millon Clinical Multiaxial Inventory (MCMI-II). *Journal of Clinical Psychology, 51,* 415–421.

Follette, W. C., & Hayes, S. C. (1992). Behavioral assessment in the *DSM* era. *Behavioral Assessment, 14,* 293–295.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state. *Journal of Psychiatric Research, 12,* 189–198.

Folstein, M. F., Romanoski, A. J., Nestadt, G., Chahal, R., Merchant, A., Shapiro, S., et al. (1985). Brief report on the clinical reappraisal of the Diagnostic Interview

Schedule carried out at the Johns Hopkins site of the Epidemiological Catchment Area Program of the NIMH. *Psychological Medicine, 15,* 809–814.

Fordyce, D. J., Roueche, J. R., & Prigatano, G. P. (1983). Enhanced emotional reactions in chronic head trauma patients. *Journal of Neurology, Neurosurgery, and Psychiatry, 46,* 620–624.

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7,* 248–260.

Frank, G. (1990). Research on the clinical usefulness of the Rorschach: The diagnosis of schizophrenia. *Perceptual and Motor Skills, 71,* 573–578.

Frank, J. D. (1973). *Persuasion and healing: A comparative study of psychotherapy* (Rev. ed.). Baltimore: Johns Hopkins University Press.

Franzen, M. D., & Iverson, G. L. (1998). Detecting negative response bias and diagnosing malingering: The dissimulation exam. In P. J. Snyder & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment* (pp. 88–101). Washington, DC: American Psychological Association.

Franzen, M. D., & Iverson, G. L. (2000). The Wechsler Memory Scales. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 195–222). New York: Wiley.

Fray, P. J., Robbins, T. W., & Sahakian, B. J. (1996). Neuropsychological applications of CANTAB. *International Journal of Geriatric Psychiatry, 11,* 329–336.

Free, M. L., & Oei, T. P. S. (1989). Biological and psychological processes in the treatment and maintenance of depression. *Clinical Psychology Review, 9,* 653–688.

Freed, E. X. (1969). Acturial data on Bender Gestalt rotations in a psychiatric hospital. *Journal of Clinical Psychology, 25,* 252–255.

Freedman, D. (2001). False prediction of future dangerousness: Error rates and Psychopathy Checklist-Revised. *Journal of the American Academy of Psychiatry and the Law, 29,* 89–95.

Friedman, A., Lewak, R., Nichols, D., & Webb, J. F. (2000). *Psychological assessment with the MMPI-2.* Mahwah, NJ: Erlbaum.

Friman, P. C., Handwerk, M. L., Smith, G. L., Larzelere, R. E., Lucas, C. P., & Shaffer, D. M. (2000). External validity of conduct and oppositional defiant disorders determined by the NIMH Diagnostic Interview Schedule for Children. *Journal of Abnormal Child Psychology, 28,* 277–286.

Frueh, B. C., & Kinder, B. N. (1994). The susceptibility of the Rorschach Inkblot Test to malingering of combat-related PTSD. *Journal of Personality Assessment, 62,* 280–298.

Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56,* 243–262.

Fuld, P. A. (1983). Psychometric differentiation of the dementias: An overview. In B. Reisberg (Ed.), *Alzheimer's disease: The standard reference* (pp. 201–210). New York: Free Press.

Fuld, P. A. (1984). Test profile of cholinergic dysfunction and of Alzheimer's-type dementia. *Journal of Clinical Neuropsychology, 6,* 380–392.

Fyer, A. J., Endicott, J., Manuzza, S., & Klein, D. F. (1985). *Schedule of Affective Disorders and Schizophrenia-Lifetime version* (modified for the study of anxiety disorders). New York: New York State Psychiatric Institute, Anxiety Disorder Clinic.

Fyer, A. J., Endicott, J., Manuzza, S., & Klein, D. F. (1995). *Schedule for Affective Disorders and Schizophrenia-Lifetime version* (modified for the study of anxiety disorders, updated for *DSM-IV; SADS-IV*). Unpublished measure, New York State Psychiatric Institute, Anxiety Genetics Unit.

Gainotti, G., & Marra, C. (1994). Some aspects of memory disorders clearly distinguish dementia of the Alzheimer's type from depressive pseudo-dementia. *Journal of Clinical and Experimental Neuropsychology, 16,* 65–78.

Gale, S. D., Johnson, S. C., Bigler, E. D., & Blatter, D. D. (1995). Nonspecific white matter degeneration following traumatic brain injury. *Journal of the International Neuropsychological Society, 1,* 17–28.

Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology, 62,* 543–549.

Gallucci, N. T. (1994). Criteria associated with clinical scales and Harris-Lingoes subscales of the Minnesota Multiphasic Personality Inventory with adolescent inpatients. *Psychological Assessment, 6,* 179–187.

Gambrill, E. D., & Richey, C. A. (1975). An assertion inventory for use in assessment and research. *Behavior Therapy, 6,* 550–561.

Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review, 4,* 641–655.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105,* 387–396.

Garb, H. N. (1992). The trained psychologist as expert witness. *Clinical Psychology Review, 12,* 451–467.

Garb, H. N. (1994a). Judgment research: Implications for clinical practice and testimony in court. *Applied and Preventive Psychology, 3,* 173–183.

Garb, H. N. (1994b). Toward a second generation of statistical prediction rules in psychodiagnostics and personality assessment. *Computers in Human Behavior, 10,* 377–394.

Garb, H. N. (1998a). Recommendations for training in the use of the Thematic Apperception Test (TAT). *Professional Psychology: Research and Practice, 29,* 621–622.

Garb, H. N. (1998b). *Studying the clinician.* Washington, DC: American Psychological Association.

Garb, H. N. (1999). Call for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings. *Assessment, 6,* 313–317.

Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that *Psychological Assessment, 12,* 31–39.

Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from a meta-analysis. *Psychological Science, 9,* 402–404.

Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Toward a resolution of the Rorschach controversy. *Psychological Assessment, 13,* 433–448.

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* New York: Basic Books.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century.* New York: Basic Books.

Garfield, S. L. (1994). Research on client variables in psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 190–228). New York: Wiley.

Garner, D. M., & Garfinkel, P. E. (1979). The Eating Attitudes Test: An index of the symptoms of anorexia nervosa. *Psychological Medicine, 9,* 273–279.

Garvey, M. J., Hollon, S. D., & DeRubeis, R. J. (1994). Do depressed patients with higher pretreatment stress levels respond better to cognitive therapy than imipramine? *Journal of Affective Disorders, 32,* 45–50.

Gass, C. S. (1991). MMPI-2 interpretation and closed-head injury: A correction factor. *Psychological Assessment, 3,* 27–31.

Gass, C. S. (1992). MMPI-2 interpretation of patients with cerebrovascular disease: A correction factor. *Archives of Clinical Neuropsychology, 7,* 17–27.

Gass, C. S. (2000). Assessment of emotional functioning with the MMPI-2. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 457–532). New York: Wiley.

Gass, C. S., & Ansley, J. (1995). Personality assessment of neurologically impaired patients. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 192–210). New York: Oxford University Press.

Gass, C. S., & Apple, C. (1997). Cognitive complaints in closed-head injury: Relationship to memory test performance and emotional disturbance. *Journal of Clinical and Experimental Neuropsychology, 19,* 290–299.

Gass, C. S., & Brown, M. C. (1992). Neuropsychological test feedback to patients with brain dysfunction. *Psychological Assessment, 4,* 272–277.

Gass, C. S., & Daniel, S. K. (1990). Emotional impact on Trail Making Test performance. *Psychological Reports, 67,* 435–438.

Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the Trail Making Test: What makes Part B harder? *Journal of Clinical and Experimental Neuropsychology, 17,* 529–535.

Gaw, K. F., & Beutler, L. E. (1995). Integrating treatment recommendations. In L. E. Beutler & M. R. Berren (Eds.), *Integrative assessment of adult personality* (pp. 280–319). New York: Guilford Press.

Geffen, G., Moar, K. J., O'Hanlon, A. P., Clark, C. R., & Geffen, L. B. (1990). Performance measures of 16- to 86-year-old males and females on the Auditory Verbal Learning Test. *Clinical Neuropsychologist, 4,* 45–63.

Geller, J., Cockell, S. J., & Drab, D. L. (2001). Assessing readiness for change in the eating disorder: The psychometric properties of the Readiness and Motivation Interview. *Psychological Assessment, 13,* 189–198.

Gerstle, R. M., Geary, D. C., Himelstein, P., & Reller-Geary, L. (1988). Rorschach predictors of therapeutic outcome for inpatient treatment of children: A proactive study. *Journal of Clinical Psychology, 44,* 277–280.

Gibertini, M., Brandenberg, N. A., & Retzlaff, P. D. (1986). The operating characteristics of the Millon Clinical Multiaxial Inventory. *Journal of Personality Assessment, 50,* 554–567.

Gieser, L., & Stein, M. I. (1999). *Evocative images: The Thematic Apperception Test and the art of projection.* Washington, DC: American Psychological Association.

Gill, D. M., Reddon, J. R., Stefanyk, W. O., & Hans, H. S. (1986). Finger tapping: Effects of trials and sessions. *Perceptual and Motor Skills, 62,* 675–678.

Gill, H. S. (1966). Delay of response and reaction to color on the Rorschach. *Journal of Projective Techniques and Personality Assessment, 30,* 545–552.

Gilley, D. W., Wilson, R. S., Fleischmann, D. A., Harrison, D. W., Goetz, C. G., & Tanner, C. M. (1995). Impact of Alzheimer's-type dementia and information source on the assessment of depression. *Psychological Assessment, 7,* 42–48.

Gilmore, D. C., Beehr, T. A., & Love, K. G. (1986). Effects of applicant sex, applicant physical attractiveness, and type of job on interview decisions. *Journal of Occupational Psychology, 59,* 103–109.

Glass, C. R., & Merluzzi, T. V. (2000). Cognitive and behavioral assessment. In C. E. Watkins & V. L. Campbell (Eds.), *Testing and assessment in counseling practice* (pp. 175–224). Mahwah, NJ: Erlbaum.

Glass, C. R., Merluzzi, T. V., Biever, J. L., & Larsen, K. H. (1982). Cognitive assessment of social anxiety: Development and validation of a self-statement questionnaire. *Cognitive Therapy and Research, 6,* 37–55.

Glutting, J. J., McDermott, P. A., Konold, T. R., Snelbaker, A. J., & Watkins, M. W. (1998). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review, 27,* 599–612.

Gocka, E. (1965). *American Lake norms for 200 MMPI scales.* Unpublished raw data.

Godber, T., Anderson, V., & Bell, R. (2000). The measurement and diagnostic utility of intrasubtest scatter in pediatric neuropsychology. *Journal of Clinical Psychology, 56,* 101–112.

Goldberg, E., & Bilder, R. M. (1987). The frontal lobes and heirarchical organization of cognitive control. In E. Perecman (Ed.), *The frontal lobes revisited* (pp. 159–187). New York: IRBN Press.

Golden, C. J. (1976). The identification of brain damage by an abbreviated form of the Halstead-Reitan Neuropsychological Battery. *Journal of Clinical Psychology, 32,* 821–826.

Golden, C. J. (1979). *Clinical interpretation of objective psychological tests.* New York: Grune & Stratton.

Golden, C. J., Purisch, A. D., & Hammeke, T. A. (1985). *Luria-Nebraska Neuropsychological Battery: Forms I and II* (Manual). Los Angeles: Western Psychological Services.

Goldfried, M. R. (1982a). Behavioral assessment: An overview. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International handbook of behavior modification and therapy.* New York: Pergamon Press.

Goldfried, M. R. (1982b). On the history of therapeutic integration. *Behavior Therapy, 13,* 572–593.

Goldfried, M. R. (1983). A behavior therapist looks at reapproachment. *Journal of Humanistic Psychology, 23,* 97–107.

Goldfried, M. R., Stricker, G., & Weiner, I. B. (1971). *Rorschach handbook of clinical and research application.* Englewood Cliffs, NJ: Prentice-Hall.

Goldman, R. D. (1960). Changes in Rorschach performance and clinical improvement in schizophrenia. *Journal of Consulting Psychology, 24,* 403–407.

Goldman, R. D., & Hartig, L. (1976). The WISC may not be a valid predictor of school performance for primary grade minority children. *American Journal of Mental Deficiency, 80,* 583–587.

Goldschmid, M. L. (1967). Prediction of college majors by personality tests. *Journal of Counseling Psychology, 14,* 302–308.

Goldstein, G., & Watson, J. R. (1989). Test-retest reliability of the Halstead-Reitan Battery and the WAIS in a neuropsychiatric population. *Clinical Neuropsychologist, 3,* 265–272.

Goldstein, J. W. (1974). Motivations for psychoactive drug use among students. In B. Kleinmuntz (Ed.), *Readings in the essentials of abnormal psychology* (pp. 371–375). New York: Harper & Row.

Goncalves, A. A., Woodward, M. J., & Millon, T. (1994). The Millon Clinical Multiaxial Inventory-II. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 161–184). Hillsdale, NJ: Erlbaum.

Goodglass, H., & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination (BDAE).* Philadelphia: Lea & Febiger. (Distributed by Psychological Assessment Resources)

Gordon, N. G., & Swart, E. C. (1973). A comparison of the Harris-Lingoes subscales between the original standardization population and an inpatient Veterans Administration hospital population. *Newsletter for Research in Mental Health and Behavioral Sciences, 15,* 28–31.

Gordon, W. A., Hibbard, M. R., Egelkos, S., & Riley, E. (1991). Issues in the diagnosis of post-stroke depression. *Rehabilitation Psychology, 36,* 71–87.

Gottschalk, L. A. (2000). The application of computerized content analysis of natural language in psychotherapy research now and in the future. *American Journal of Psychotherapy, 54,* 305–311.

Gough, H. G. (1948). A new dimension of status. I: Development of a personality scale. *American Sociological Review, 13,* 401–409.

Gough, H. G. (1957). *California Psychological Inventory manual.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1964). Academic achievement in high school as predicted from the California Psychological Inventory. *Journal of Educational Psychology, 65,* 174–180.

Gough, H. G. (1966). Graduation from high school as predicted from the California Psychological Inventory. *Psychology in the Schools, 3,* 208–216.

Gough, H. G. (1968). An interpreter's syllabus for the California Psychological Inventory. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 1, pp. 55–79). Palo Alto, CA: Science and Behavior Books.

Gough, H. G. (1969). A leadership index on the California Psychological Inventory. *Journal of Counseling Psychology, 16,* 285–289.

Gough, H. G. (1975). *Manual for the California Psychological Inventory* (Rev. ed.). Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1987). *California Psychological Inventory: Administrator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1992). Assessment of creative potential in psychology and the development of a creative temperament scale for the CPI. In J. C. Rusen & P. McReynolds (Eds.), *Advances in psychological assessment* (Vol. 8, pp. 225–257). New York: Plenum Press.

Gough, H. G. (1996). *California Psychological Inventory manual.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (2000). The California Psychological Inventory. In C. E. Watkins & V. L. Campbell (Eds.), *Testing and assessment in counseling practice* (2nd ed., pp. 45–71). Mahwah, NJ: Erlbaum.

Gough, H. G., & Bradley, P. (1992). Delinquent and criminal behavior as assessed by the revised California psychological Inventory. *Journal of Clinical Psychology, 48,* 298–308.

Gough, H. G., Durflinger, G. W., & Hill, R. E., Jr. (1968). Predicting performance in student teaching from the California Psychological Inventory. *Journal of Educational Psychology, 52,* 119–127.

Gough, H. G., & Hall, W. B. (1964). Prediction of performance in medical school from the California Psychological Inventory. *Journal of Applied Psychology, 48,* 218–226.

Gough, H. G., & Kirk, B. A. (1970). Achievement in dental school as related to personality and aptitude variables. *Measurement and Evaluation in Guidance, 2,* 225–233.

Gough, H. G., & Lanning, K. (1986). Predicting grades in college from the California Psychological Inventory. *Educational and Psychological Measurement, 46,* 205–213.

Gough, H. G., Wenk, E. A., & Rozynko, V. V. (1965). Parole outcome as predicted from the CPI, the MMPI, and a Base Expectancy Table. *Journal of Abnormal Psychology, 70,* 432–441.

Gould, R. A., Otto, M. W., & Pollack, M. H. (1995). A meta-analysis of treatment outcome for panic disorder. *Clinical Psychology Review, 15,* 819–844.

Grace, J., Stout, J., & Malloy, P. F. (1999). Assessing frontal lobe behavioral syndromes with the Frontal Lobe Personality Scale. *Assessment, 6,* 269–284.

Graham, J. R. (1978). A review of some important MMPI special scales. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 4, pp. 311–331). San Francisco: Jossey-Bass.

Graham, J. R. (1987). *The MMPI: A practical guide* (2nd ed.). New York: Oxford University Press.

Graham, J. R. (2000). *MMPI-2: Assessing personality and psychopathology* (3rd ed.). New York: Oxford University Press.

Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). *MMPI-2 correlates for outpatient mental health settings.* Minneapolis: University of Minnesota Press.

Graham, J. R., & Lilly, R. S. (1984). *Psychological testing.* Englewood Cliffs, NJ: Prentice-Hall.

Graham, J. R., & McCord, G. (1985). Interpretation of moderately elevated MMPI scores for normal subjects. *Journal of Personality Assessment, 49,* 477–484.

Graham, J. R., Smith, R. L., & Schwartz, G. F. (1986). Stability of MMPI configurations for psychiatric inpatients. *Journal of Consulting and Clinical Psychology, 49,* 477–484.

Graham, P., & Rutter, M. (1968). The reliability and validity of the psychiatric assessment of the child. II: Interview with the parent. *British Journal of Psychiatry, 114,* 581–592.

Graves, R. E., Carswell, L. M., & Snow, W. G. (1999). An evaluation of the sensitivity of premorbid IQ estimators for detecting cognitive decline. *Psychological Assessment, 11,* 29–38.

Green, S. B., & Kelley, C. K. (1988). Racial bias in prediction with the MMPI for a juvenile delinquent population. *Journal of Personality Assessment, 52,* 263–275.

Greenbaum, P. F., Prange, M. E., Friedman, R. M., & Silver, S. E. (1991). Substance abuse prevalence and comorbidity with other psychiatric disorders among adolescents with severe emotional disorders. *Journal of the Academy of Child and Adolescent Psychology, 30,* 575–583.

Greene, R. L. (1987). Ethnicity and MMPI performance: A review. *Journal of Consulting and Clinical Psychology, 55,* 497–512.

Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual.* Boston: Allyn & Bacon.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.

Greene, R. L., & Clopton, J. R. (1994). Minnesota Multiphasic Personality Inventory-2. In M. E. Maruish (Ed.), *The use of psychological testing for treatment, planning and outcome assessment* (pp. 137–159). Hillsdale, NJ: Erlbaum.

Greene, R. L., & Clopton, J. R. (1999). Minnesota Multiphasic Personality Inventory-2 (MMPI-2). In M. E. Maruish (Ed.), *The use of psychological testing for treatment, planning and outcomes assessment* (2nd ed., pp. 1023–1049). Mahwah, NJ: Erlbaum.

Greene, R. L., Weed, N. C., Butcher, J. N., & Arredondo, R. (1992). A cross validation of MMPI-2 substance abuse scales. *Journal of Personality Assessment, 58,* 405–410.

Greenspan, S., & Driscoll, J. (1997). The role of intelligence in a broad model of personal competence. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 131–150). New York: Guilford Press.

Greenwald, D. F. (1991). Personality dimensions reflected by the Rorschach and the 16PF. *Journal of Clinical Psychology, 47,* 708–715.

Greenway, P., & Milne, L. (1999). Relationship between psychopathology, learning disabilities, or both and WISC-III subtest scatter in adolescents. *Psychology in the Schools, 36,* 103–108.

Gregory, M. L. (1977). Emotional indicators on the Bender Gestalt and the Devereux Child Rating Scale. *Psychology in the Schools, 14,* 433–437.

Gregory, R. J. (1999). *Foundations of intellectual assessment: The WAIS-III and other tests in clinical practice.* Boston: Allyn & Bacon.

Greif, E. B., & Hogan, R. (1973). The theory and measurement of empathy. *Journal of Counseling Psychology, 20,* 280–284.

Gresham, F. M. (1984). Behavioral interviews in school psychology: Issues in psychometric adequacy and research. *School Psychology Review, 13,* 17–25.

Grice, J. W., Krohn, E. J., & Logerquist, S. (1999). Cross-validation of the WISC-III factor structure in two samples of children with learning disabilities. *Journal of Psychoeducational Assessment, 17,* 236–248.

Griffin, M. L., & Flaherty, M. R. (1964). Correlation of CPI traits with academic achievement. *Educational and Psychological Measurement, 24,* 369–372.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124,* 75–111.

Grillo, J., Brown, R. S., Hilsabeck, R., Price, J. R., & Lees-Haley, P. (1994). Raising doubts about claims of malingering: Implications of relationships between MCMI-II and MMPI-2 performances. *Journal of Clinical Psychology, 50,* 651–655.

Groth-Marnat, G. (1985). Evaluating and using psychological testing software. *Human Resource Management Australia, 23,* 16–21.

Groth-Marnat, G. (1988). A survey of the current and future direction of professional psychology in acute general hospitals in Australia. *Australian Psychologist, 23,* 39–43.

Groth-Marnat, G. (1991). Hypnotizability, suggestibility, and psychopathology: An overview of research. In J. F. Schumaker (Ed.), *Human suggestibility: Advances in theory, research, and application* (pp. 219–234). New York: Routledge.

Groth-Marnat, G. (1992). Past cultural traditions of therapeutic metaphor. *Psychology: An International Journal of Human Behavior, 29,* 1–8.

Groth-Marnat, G. (1993). Neuropsychological effects of styrene exposure: A review of current literature. *Perceptual and Motor Skills, 77,* 1139–1149.

Groth-Marnat, G. (1995). *The Rey Auditory Verbal Learning Test: A manual for administration, scoring, and interpretation* (Tech. and Res. Rep.). Perth, Australia: Curtin University of Technology, Research Center for Applied Psychology.

Groth-Marnat, G. (1999). Financial efficacy of clinical assessment: Rational guidelines and issues for future research. *Journal of Clinical Psychology, 55,* 813–824.

Groth-Marnat, G. (Ed.). (2000a). *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration.* New York: Wiley.

Groth-Marnat, G. (2000b). Visions of clinical assessment: Then, now, and a brief history of the future. *Journal of Clinical Psychology, 56,* 349–365.

Groth-Marnat, G. (in press). The psychological report. In R. Fernandez-Ballesteros (Ed.), *Encyclopedia of psychological assessment.* Thousand Oaks, CA: Sage.

Groth-Marnat, G., & Edkins, G. (1996). Professional psychologists in general health care settings: A review of the financial efficacy of direct treatment interventions. *Professional Psychology: Research and Practice, 27,* 161–174.

Groth-Marnat, G., Edkins, G., & Schumaker, J. F. (1995). Psychologists in disease prevention and health promotion: A review of the cost-effectiveness literature. *Psychology, 32,* 127–135.

Groth-Marnat, G., Gallagher, R. E., Hale, J. B., & Kaplan, E. (2000). The Wechsler intelligence scales. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A practical guide to test interpretation and integration* (pp. 129–194). New York: Wiley.

Groth-Marnat, G., & Roberts, L. (1998). Indicators of psychological health on Human Figure Drawings and the House Tree Person: A search for concurrent validity. *Journal of Clinical Psychology, 46,* 219–222.

Groth-Marnat, G., Roberts, R., & Beutler, L. E. (2001). Client characteristics and psychotherapy: Perspectives, support, interactions, and implications. *Australian Psychologist, 36,* 115–121.

Groth-Marnat, G., & Schumaker, J. (1989). Computer-based psychological testing: Issues and guidelines. *American Journal of Orthopsychiatry, 59,* 257–263.

Groth-Marnat, G., & Teal, M. (2000). Block Design as a measure of everyday spatial ability: A study of ecological validity. *Perceptual and Motor Skills, 90,* 522–526.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A metanalysis. *Psychological Assessment, 12,* 19–30.

Guevremont, G. C., & Spiegler, M. D. (1990, November). *What do behavior therapists really do? A survey of the clinical practice of AABT members.* Paper presented at the 24th annual convention of the Association for Advancement of Behavior Therapy, San Francisco.

Gynther, M. D. (1978). The California Psychological Inventory: A review. In O. K. Buros (Ed.), *The eighth mental measurements yearbook* (Vol. 1, pp. 733–737). Highland Park, NJ: Gryphon Press.

Gynther, M. D. (1979). Aging and personality. In J. N. Butcher (Ed.), *New developments in the use of the MMPI* (pp. 240–256). Minneapolis: University of Minnesota Press.

Gynther, M. D., & Green, S. B. (1980). Accuracy may make a difference, but does a difference make for accuracy? A response to Pritchard and Rosenblatt. *Journal of Consulting and Clinical Psychology, 48,* 268–272.

Gynther, M. D., & Shimkuras, A. M. (1966). Age and MMPI performance. *Journal of Consulting Psychology, 30,* 118–121.

Haaga, D. A., Dyck, M. J., & Ernst, D. (1991). Empirical status of cognitive theory of depression. *Psychological Bulletin, 110,* 215–236.

Haan, N. (1964). An investigation of the relationships of Rorschach scores, patterns and behaviors to coping and defense mechanisms. *Journal of Projective Techniques and Personality Assessment, 28,* 429–441.

Haas, A. P., Hendin, H., & Singer, P. (1987). Psychodynamic and structured interviewing: Issues of validity. *Comprehensive Psychiatry, 28,* 40–53.

Haddad, F. A. (1986). The performance of learning disabled children on the Kaufman Assessment Battery for Children and the Bender-Gestalt Test. *Psychology in the Schools, 23,* 342–345.

Hafner, A. J. (1958). Response time and Rorschach behavior. *Journal of Clinical Psychology, 14,* 154–155.

Hain, J. D. (1964). The Bender Gestalt Test: A scoring method for identifying brain damage. *Journal of Consulting Psychology, 28,* 34–40.

Hale, W. D., Cochran, C. D., & Hedgepeth, B. E. (1984). Norms for the elderly on the Brief Symptom Inventory. *Journal of Counsulting and Clinical Psychology, 52,* 321–322.

Hall, C. (1983, December 6). Psychiatrist's computer use stirs debate. *The Wall Street Journal, 35,* 39.

Hall, G. C. N., Bansal, A., & Lopez, I. R. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment, 11,* 186–197.

Hall, H. V., & Sbordone, R. J. (Eds.). (1993). *Disorders of executive function.* Winter Park, FL: PMD Publishers.

Hammen, C. L. (1978). Depression, distortion, and life stress in college students. *Cognitive Therapy and Research, 2,* 189–192.

Hammen, C. L., & Krantz, S. (1976). Effect of success and failure on depressive cognitions. *Journal of Abnormal Psychology, 85,* 577–586.

Handel, R. W., & Ben-Porath, Y. S. (2000). Multicultural assessment with the MMPI-2: Issues for research and practice. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 229–245). Mahwah, NJ: Erlbaum.

Handelsman, M. M., & Galvin, M. D. (1988). Facilitating informed consent for outpatient psychotherapy: A suggested written format. *Professional Psychology: Research and Practice, 19,* 223–225.

Hansell, A. G., Lerner, H. D., Milden, R. S., & Ludolph, P. (1988). Single-sign Rorschach suicide indicators: A validity study using a depressed inpatient population. *Journal of Personality Assessment, 52,* 658–669.

Hanson, R. K., Hunsley, J., & Parker, K. C. H. (1988). The relationship between WAIS subtest reliability, "g" loadings, and meta-analytically derived validity estimates. *Journal of Clinical Psychology, 44,* 557–562.

Hargrave, G. E., Hiatt, D., & Gaffney, T. W. (1986). A comparison of MMPI and CPI test profiles for traffic officers and deputy sheriffs. *Journal of Police Science and Administration, 14,* 250–258.

Hargrave, G. E., Hiatt, D., Ogard, E. M., & Karr, C. (1994). Comparison of the MMPI and MMPI-2 for a sample of peace officers. *Psychological Assessment, 6,* 27–32.

Harrell, T. H., Honaker, L. M., & Parnell, T. (1992). Equivalence of the MMPI-2 with the MMPI in psychiatric patients. *Psychological Assessment, 4,* 460–465.

Harris, R. E., & Lingoes, J. C. (1968). *Subscales for the MMPI: An aid to profile interpretation.* Los Angeles: University of California, Department of Psychiatry. (Original work published 1955)

Harrison, R. (1940). Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients. II: A quantitative validity study. III: Validation by blind analysis. *Character and Personality, 9,* 122–133, 134–138.

Harrower, M. (1986). The Stress Tolerance Test. *Journal of Personality Assessment, 50,* 417–427.

Hart, R. P., Kwentus, J. A., Taylor, J. R., & Hamer, R. M. (1988). Productive naming and memory in depression and Alzheimer's type dementia. *Archives of Clinical Neuropsychology, 3,* 313–322.

Hartshorne, H., & May, M. A. (1928). *Studies in deceit.* New York: Macmillan.

Harvey, V. S. (1997). Improving readability of psychological reports. *Professional Psychology: Research and Practice, 28,* 271–274.

Harwood, T. M., & Williams, O. (2003). Identifying treatment relevant assessment: The STS. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality.* New York: Guilford Press.

Hase, H. D., & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin, 67,* 231–248.

Hathaway, S. R., & McKinley, J. C. (1940). A Multiphasic Personality Schedule (Minnesota). I: Construction of the schedule. *Journal of Psychology, 10,* 249–254.

Hathaway, S. R., & Monachesi, E. D. (1963). *Adolescent personality and behavior: MMPI patterns of normal, delinquent, dropout, and other outcomes.* Minneapolis: University of Minnesota Press.

Haverkamp, B. E. (1993). Confirmatory bias in hypothesis testing for client-identified and counselor self-generated hypotheses. *Journal of Counseling Psychology, 40,* 303–315.

Hawkins, K. A. (1998). Indicators of brain dysfunction derived from graphic representations of the WAIS-III/WMS-III Technical manual clinical sample data: A preliminary approach to clinical utility. *Clinical Neuropsychologist, 12,* 535–551.

Hawkins, K. A., Sullivan, T. E., & Choi, E. J. (1997). Memory defects in schizophrenia: Inadequate assimilation or true amnesia? Findings from the Wechsler Memory Scale-Revised. *Journal of Psychiatry and Neuroscience, 22,* 169–179.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311–327.

Hayes, J. A. (1997). What does the Brief Symptom Inventory measure in college and university counseling center clients? *Journal of Counseling Psychology, 44,* 360–367.

Haynes, J. P., & Howard, R. C. (1986). Stability of WISC-R scores in a juvenile forensic sample. *Journal of Clinical Psychology, 42,* 534–537.

Haynes, J. P., & Peltier, J. (1985). Patterns of practice with the TAT in juvenile forensic settings. *Journal of Personality Assessment, 49,* 20–29.

Haynes, S. N. (1991). Clinical applications of psychophysiological assessment: An introduction and overview. *Psychological Assessment, 3,* 307–308.

Haynes, S. N. (1995). Introduction to the special section in chaos theory and psychological assessment. *Psychological Assessment, 7,* 3–4.

Haynes, S. N., & O'Brien, W. H. (1988). The Gordian knot of *DSM-III-R* use: Integrating principles of behavior classification and complex causal models. *Behavioral Assessment, 10,* 95–106.

Haynes, S. N., & O'Brien, W. H. (2000). *Principles and practice of behavioral assessment.* New York: Kluwer Academic.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7,* 238–247.

Heaton, R. K., Beade, L. E., & Johnson, K. L. (1978). Neuropsychological test results associated with psychiatric disorders in adults. *Psychological Bulletin, 85,* 141–162.

Heaton, R. K., Grant, I., & Mathews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan Battery.* Odessa, FL: Psychological Assessment Resources.

Hebb, D. O. (1972). *Textbook of psychology* (3rd ed.). Philadelphia: Saunders.

Hebben, N., & Milberg, W. (2002). *Essentials of neuropsychological assessment.* New York: Wiley.

Hedlund, J. L., Sletten, I. W., Evenson, R. C., Altman, H., & Cho, D. W. (1977). Automated psychiatric information systems: A critical review of Missouri's Standard System of Psychiatry (SSOP). *Journal of Operational Psychiatry, 8,* 5–26.

Heflinger, C. A., Cook, V. J., & Thackrey, M. (1987). Identification of mental retardation by the System of Multicultural Pluralistic Assessment: Nondiscriminatory or nonexistent? *Journal of School Psychology, 25,* 177–183.

Heilbrun, A. B., Daniel, J. L., Goodstein, L. D., Stephenson, R. R., & Crites, J. O. (1962). The validity of two-scale pattern interpretation on the California Psychological Inventory. *Journal of Applied Psychology, 46,* 409–416.

Heilburn, K. (2001). *Principles of forensic mental health assessment.* New York: Kluwer Academic/Plenum Press.

Heimberg, R. G. (1994). Cognitive assessment strategies and the measurement of outcome of treatment for social phobia. *Behavior Research and Therapy, 32,* 269–280.

Heimberg, R. G., Harrison, D. F., Goldberg, L. S., Desmarais, S., & Blue, S. (1979). The relationship of self-report and behavioral assertion in an offender population. *Journal of Behavior Therapy and Experimental Psychiatry, 10,* 283–286.

Hellkamp, D. T., & Hogan, M. E. (1985). Differentiation of organics from functional psychiatric patients across various I.Q. ranges using the Bender-Gestalt and Hutt scoring system. *Journal of Clinical Psychology, 41,* 259–264.

Helmes, E. (2000). Learning and memory. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 293–334). New York: Wiley.

Helmes, E., & Reddon, J. R. (1993). A perspective on development in assessing psychopathology: A critical review of MMPI and MMPI-2. *Psychological Bulletin, 113,* 453–471.

Helzer, J. E., & Robins, L. N. (1988). The Diagnostic Interview Schedule: Its development, evolution and use. *Social Psychiatry and Psychiatric Epidemiology, 23,* 6–16.

Helzer, J. E., Robins, L. N., Croughan, J. L., & Welner, A. (1981). Renard Diagnostic Interview: Its reliability and procedural validity with physicians and lay interviewers. *Archives of General Psychiatry, 38,* 393–398.

Helzer, J. E., Robins, L. N., McEvoy, L. F., Spitznagel, E. L., Stolzman, R. K., Farmer, A., et al. (1985). A comparison of Clinical and Diagnostic Interview Schedule diagnoses: Physician re-examination of lay-interviewed cases in the general population. *Archives of General Psychiatry, 42,* 657–666.

Helzer, J. E., Spitznagel, E. L., & McEvoy, L. (1987). The predictive validity of lay Diagnostic Interview Schedule diagnoses in the general population: A comparison with physician examiners. *Archives of General Psychiatry, 44,* 1069–1077.

Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the "remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment, 6,* 92–101.

Henry, E. M., & Rotter, J. B. (1956). Situational influences on Rorschach responses. *Journal of Consulting Psychology, 20,* 457–462.

Henry, W. E. (1956). *The analysis of fantasy: The Thematic Apperception Test in the study of personality.* New York: Wiley.

Herjanic, B., & Campbell, W. (1977). Differentiating psychiatrically disturbed children on the basis of a structured interview. *Journal of Abnormal Child Psychology, 51,* 127–134.

Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement on diagnosis comparing child and patient interviews. *Journal of Abnormal Psychology, 10,* 325–336.

Hersen, M. (1988). Behavioral assessment and psychiatric diagnosis. *Behavioral Assessment, 10,* 107–121.

Hersen, M., & Bellack, A. S. (1976). *Behavioral assessment: A practical handbook.* New York: Pergamon Press.

Hersen, M., & Bellack, A. S. (1988). *DSM-III* and behavioral assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3rd ed., pp. 67–84). New York: Pergamon Press.

Hersen, M., & Bellack, A. S. (Eds.). (1998). *Dictionary of behavioral assessment techniques* (2nd ed.). New York: Pergamon Press.

Hersen, M., & Greaves, S. T. (1971). Rorschach productivity as related to verbal performance. *Journal of Personality Assessment, 35,* 436–441.

Hertz, M. R. (1948). Suicidal configurations in Rorschach records. *Rorschach Research Exchange, 12,* 3–58.

Hertz, M. R. (1976). Detection of suicidal risks with the Rorschach. In M. Abt & S. L. Weissman (Eds.), *Acting out: Theoretical and clinical aspects* (2nd ed.). New York: Aronson.

Herzog, D. B., Keller, M. B., Sacks, N. R., Yeh, C. J., & Lavori, P. W. (1992). Psychiatric comorbidity in treatment-seeking anorexics and bulimics. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 810–818.

Higgins, R. L., Alonso, R. R., & Pendleton, M. G. (1979). The validity of roleplay assessments of assertiveness. *Behavior Therapy, 10,* 655–662.

Higgins-Lee, C. (1990). Low scores on California Psychological Inventory as predictors of psychopathology in alcoholic patients. *Psychological Reports, 67,* 227–232.

Hill, A. H. (1967). Use of a structured autobiography in the construct validation of personality scales. *Journal of Consulting Psychology, 31,* 551–556.

Hill, R. E., Jr. (1960). Dichotomous prediction of student teaching excellence employing selected CPI scales. *Journal of Educational Research, 53,* 349–351.

Hilsenroth, M. J., & Handler, L. (1995). A survey of graduate students' experiences, interests, and attitudes about learning the Rorschach. *Journal of Personality Assessment, 64,* 243–257.

Hishinuma, E. S., & Yamakawa, R. (1993). Constructional and criterion related validity of the WISC-III for exceptional students and those at risk. In B. A. Braken & R. S. McCullum (Eds.), *Wechsler Intelligence Scales for Children* (3rd ed.). Brandon, VT: Clinical Psychology.

Hogan, R. (1971). Personality characteristics of highly rated policemen. *Personnel Psychology, 24,* 679–686.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51,* 469–477.

Hoge, R. D., Andrews, D. A., Robinson, D., & Hollett, J. (1988). The construct validity of interview-based assessments in family counseling. *Journal of Clinical Psychology, 44,* 563–571.

Holland, A. L. (1980). *Communicative abilities in daily living: A test of functional communication for aphasic adults.* Austin, TX: ProEd.

Holliman, N. B., & Montross, J. (1984). The effects of depression upon responses to the California Psychological Inventory. *Journal of Clinical Psychology, 40,* 1373–1378.

Hollis, J. W., & Donna, P. A. (1979). *Psychological report writing: Theory and practice.* Muncie, IN: Accelerated Development.

Hollon, S. D., & Kendall, P. C. (1980). Cognitive self-statements in depression: Development of an automatic thoughts questionnaire. *Cognitive Therapy and Research, 4,* 383–395.

Holmes, C. B., Dungan, D. S., & Medlin, W. J. (1984). Reassessment of inferring personality traits from Bender-Gestalt drawing styles. *Journal of Clinical Psychology, 40,* 1241–1243.

Holmes, C. B., & Stephens, C. L. (1984). Consistency of edging on the Bender-Gestalt, Memory for Designs, and Draw-A-Person Tests. *Journal of Psychology, 117,* 269–271.

Holmstrom, R. W., Karp, S. A., & Silber, D. E. (1994). Prediction of depression with the Apperceptive Personality Test. *Journal of Clinical Psychology, 50,* 234–237.

Holtzman, W. H. (1988). Beyond the Rorschach. *Journal of Personality Assessment, 52,* 578–609.

Hope, D. A., & Heimberg, R. G. (1993). Social phobia and social anxiety. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (pp. 99–136). New York: Guilford Press.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. Wolman (Ed.), *Handbook of intelligence* (pp. 267–300). New York: Wiley.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystalized intelligence. *Journal of Educational Psychology, 57,* 253–270.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues.* New York: Guilford Press.

Horowitz, M. J. (1985). *Report of the program on conscious and unconscious mental processes of the John D. and Catherine T. MacArthur Foundation.* San Francisco: University of California.

Horvath, A. O., & Goheen, M. D. (1990). Factors mediating the success of defiance- and compliance-based interventions. *Journal of Consulting and Clinical Psychology, 37,* 363–371.

Horvath, A. O., & Symonds, B. D. (1991). Relations between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38,* 139–149.

Houts, A. C. (2002). Discovery, invention, and the expansion of the modern *Diagnostic and Statistical Manuals of Mental Disorders.* In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 17–69). Washington, DC: American Psychological Association.

Hoyt, M. F. (1994). Single session solutions. In M. Hoyt (Ed.), *Constructive therapies* (pp. 140–159). New York: Guilford Press.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter. (1984). revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79,* 184–194.

Humphrey, D. H., & Dahlstrom, W. G. (1995). The impact of changing from the MMPI to the MMPI-2 on profile configurations. *Journal of Personality Assessment, 64,* 428–439.

Hunkin, N. M., Stone, J. V., Isaac, C., Holdstock, J. S., Butterfield, R., Wallis, L. I., et al. (2000). Factor analysis of three standardized tests of memory in a clinical population. *British Journal of Clinical Psychology, 39,* 169–180.

Hunsley, J., & Bailey, J. M. (2001). Whither the Rorschach? *Psychological Assessment, 13,* 472–485.

Hunsley, J., Hanson, R. K., & Parker, K. C. H. (1988). A summary of the reliability and stability of MMPI scales. *Journal of Clinical Psychology, 44,* 44–46.

Hunter, E. E., Powell, B. J., Penick, E. C., Nickel, E. J., Othmer, E., & DeSouza, C. (2000). Development and validation of a mania scale for the Symptom Checklist-90-R. *Journal of Nervous and Mental Diseases, 188,* 176–179.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340–362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job employment. *Psychological Bulletin, 96,* 72–98.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2,* 447–472.

Hutt, M. L. (1985). *The Hutt adaptation of the Bender-Gestalt Test* (4th ed.). New York: Grune & Stratton.

Hutt, M. L., & Briskin, G. J. (1960). *The clinical use of the revised Bender-Gestalt Test.* New York: Grune & Stratton.

Iacono, W. G. (1991). Psychophysiological assessment of psychopathology. *Psychological Assessment, 3,* 309–320.

Impara, J. C., & Plake, B. S. (1998). *Thirteenth Mental Measurements Yearbook*. Lincoln, NE: Buros Institute.

Ingram, R. E., Kendall, P. C., Siegle, G., Guarino, J., & McLaughlin, S. C. (1995). Psychometric properties of the Positive Automatic Thoughts Questionnaire. *Psychological Assessment, 7,* 495–507.

Insua, A. M., & Stella, M. (1986). Psychometric patterns on the Rorschach of healthy elderly persons and patients with suspected dementia. *Perceptual and Motor Skills, 63,* 931–936.

Ireland-Galman, M., Padilla, G., & Michael, W. (1980). The relationship between performance on the mazes subtest of the Wechsler Intelligence Scale for Children-Revised (WISC-R) and speed of solving anagrams with simple and difficult arrangements of letter and order. *Educational and Psychological Measurement, 40,* 513–524.

Iverson, G. L. (1999). Interpreting change on the WAIS-III/WMS-III following traumatic brain injury. *Journal of Cognitive Rehabilitation,* 16–20.

Iverson, G. L., Franzen, M. D., & Hammond, J. A. (1995). Examination of inmate's ability to malinger on the MMPI-2. *Psychological Assessment, 7,* 115–117.

Ivnik, R. J., Malec, J. F., Tangalos, E. G., Peterson, R. C., Kokmen, E., & Kurland, L. (1990). The Auditory-Verbal Learning Test (AVLT): Norms for ages 55 years and older. *Psychological Assessment, 2,* 304–312.

Ivnik, R. J., Sharbrough, F. W., & Laws, E. R. (1988). Anterior temporal lobectomy for the control of partial complex seizures: Information for counseling patients. *Mayo Clinic Proceedings, 63,* 783–793.

Jackson, D. N., Fraboni, M., & Helms, E. (1997). MMPI-2 content scales: How much content do they measure? *Assessment, 4,* 111–117.

Jackson, M., Fox, G. A., Waugh, M., & Tuck, R. R. (1987). The Rey Auditory Verbal Learning Test: A follow-up study on alcoholics with and without neurological disease. *Australian Drug and Alcohol Review, 6,* 71–76.

Jacob, R. G., Turner, S. M., Szekely, B. C., & Eidelman, B. H. (1983). Predicting outcome of relaxation therapy in headaches: The role of "depression." *Behavior Therapy, 14,* 457–465.

Jacobson, J. W., & Mulick, J. A. (Eds.). (1996). *Manual of diagnosis and professional practice in mental retardation.* Washington, DC: American Psychological Association.

Jacobson, N. S., Follette, W. C., & Pagel, M. (1986). Predicting who will benefit from behavioral marital therapy. *Journal of Consulting and Clinical Psychology, 54,* 518–522.

Jankowski, D. (2002). *A beginner's guide to the MCMI–III.* Washington, DC: American Psychological Association.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia, 27,* 1043–1056.

Janus, M., Tolbert, H., Calestro, K., & Toepfer, S. (1996). Clinical accuracy ratings of MMPI approaches for adolescents: Adding ten years and the MMPI-A. *Journal of Personality Assessment, 67,* 364–383.

Jarvis, P. E., & Barth, J. (1994). *The Halstead-Reitan Neuropsychological Battery: A guide to interpretation and clinical application.* Odessa, FL: Psychological Assessment Resources.

Jastak, S., & Wilkinson, G. (1994). *Wide Range Achievement Test* (3rd ed.). San Antonio, TX: Psychological Corporation.

Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnosis for clinireferred children and adolescents. *Journal of Consulting and Clinical Psychology, 70,* 158–168.

Jensen, A. R. (1965). Review of the Rorschach. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (pp. 501–509). Highland Park, NJ: Gryphon Press.

Jensen, A. R. (1969). How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review, 39,* 1–23.

Jensen, A. R. (1972). *Genetics and education.* New York: Harper & Row.

Jensen, A. R. (1984). The black-white difference on the K-ABC: Implications for future tests. *Journal of Special Education, 18,* 377–408.

Jensen, A. R., & Reynolds, C. R. (1982). Race, social class, and ability patterns on the WISC-R. *Personality and Individual Differences, 3,* 423–438.

Jensen, A. R., & Rohwer, W. D. (1966). The Stroop Color-Word Test: A review. *Acta Psychologica, 25,* 36–93.

Johnson, D. L., & Danley, W. (1981). Validity: Comparison of the WISC-R and SOMPA estimated learning potential scores. *Psychological Reports, 49,* 123–131.

Johnson, J. A. (2000). Predicting observers' ratings of the Big Five from the CPI, HPI, and NEO-PI-R: A comparative validity study. *European Journal of Personality, 14,* 1–19.

Johnson, J. H., Null, C., Butcher, J. M., & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology, 47,* 105–114.

Johnson, J. W., & Williams, T. A. (1977). Using on-line computer technology to improve service response and decision-making effectiveness in a mental health admitting system. In J. B. Sidowski & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 237–249). Norwood, NJ: Ablex.

Johnson, L. C., Murphy, S. A., Dimond, M. (1996). Reliability, construct validity, and subscale norms of the Brief Symptom Inventory when administered to bereaved parents. *Journal of Nursing Measurement, 4,* 117–127.

Johnson, M. H., Margo, P. A., & Stern, S. L. (1986). Use of the SADS-C as a diagnostic and symptom severity measure. *Journal of Consulting and Clinical Psychology, 54,* 546–551.

Jones, M. C. (1924). The elimination of children's fears. *Journal of Experimental Psychology, 7,* 382–390.

Jones, R. G. (1969). A factored measure of Ellis's Irrational Belief System. *Dissertation Abstracts International, 29*(43), 4379B–4380B. (UMI No. 69–64)

Jongsma, A. E., & Peterson, L. M. (1995). *The complete psychotherapy treatment planner.* New York: Wiley.

Jorgensen, K., Anderson, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: A review. *Assessment, 7,* 259–280.

Julian, A., Sotile, W. M., Henry, S. H., & Sotile, M. O. (1991). *Technical manual: The Family Apperception Test.* Beverly Hills, CA: Western Psychological Services.

Kabacoff, R. I., Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of Anxiety Disorders, 11,* 33–47.

Kadden, R. M., Cooney, N. L., Getter, H., & Litt, M. D. (1990). Matching alcoholics to coping skills or interactional therapies: Post-treatment results. *Journal of Consulting and Clinical Psychology, 57,* 698–704.

Kadden, R. M., Litt, M. D., Donovan, D., & Cooney, N. L. (1996). Psychometric properties of the California Psychological Inventory Socialization Scale in treatment-seeking alcoholics. *Psychology of Addictive Behaviors, 10,* 131–146.

Kahn, M. W., Fox, H., & Rhode, R. (1988). Detecting faking on the Rorschach: Computer versus expert clinical judgment. *Journal of Personality Assessment, 52,* 516–523.

Kahn, R. L., & Cannell, C. F. (1961). *The dynamics of interviewing: Theory, technique, and cases.* New York: Wiley.

Kahn, T. C., & Giffen, M. B. (1960). *Psychological techniques in diagnosis and evaluation.* New York: Pergamon Press.

Kallingal, A. (1971). The prediction of grades for black and white students at Michigan State University. *Journal of Educational Measurement, 8,* 263–265.

Kamin, L. J. (1974). *The science and politics of I.Q.* Hillsdale, NJ: Erlbaum.

Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice, 31,* 155–164.

Kane, R. L. (1991). Standardized and flexible batteries in neuropsychology: An assessment update. *Neuropsychology Review, 2,* 281–337.

Kane, R. L., & Kay, G. G. (1992). Computerized assessment in neuropsychology: A review of tests and test batteries. *Neuropsychology Review, 3,* 1–117.

Kanfer, F. H., & Grimm, L. G. (1977). Behavioral analysis: Selecting target behaviors in the interview. *Behavior Modification, 4,* 419–444.

Kanfer, F. H., & Saslow, G. (1969). Behavioral diagnosis. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status.* New York: McGraw-Hill.

Kaplan, E., Fein, D., Morris, R., & Delis, D. (1991). *WAIS-R as a neuropsychological instrument.* San Antonio, TX: Psychological Corporation.

Kaplan, E., Fein, D., Morris, R., Kramer, J. H., & Delis, D. C. (1999). *The WISC-III as a process instrument.* San Antonio, TX: Psychological Corporation.

Kaplan, H. I., & Sadock, B. J. (2001). *Kaplan & Sadock's pocket handbook of clinical psychiatry* (3rd ed.). Philadelphia: Lippincott.

Kaplan, R. M., & Saccuzzo, D. P. (1993). *Psychological testing: Principles, applications and issues* (3rd ed.). Belmont, CA: Wadsworth.

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues* (5th ed.). Pacific Grove, CA: Brooks/Cole.

Kareken, D. A., & Williams, J. M. (1994). Human judgment and estimation of premorbid intellectual function. *Psychological Assessment, 6,* 83–91.

Karon, B. P. (2000). The clinical interpretation of the Thematic Apperception Test, Rorschach, and other clinical data: A reexamination of statistical versus clinical prediction. *Professional Psychology: Research and Practice, 31,* 230–233.

Karp, S. A., Holstrom, R. W., & Silber, D. E. (1989). *Manual for the Apperceptive Personality Test (APT).* Orland Park, IL: International Diagnostic Systems.

Karp, S. A., Silber, D. E., Holstrom, R. W., Banks, V., & Karp, J. (1992). Outcomes of Thematic Apperceptive Test and Apperceptive Personality Test stories. *Perceptual and Motor Skills, 74,* 479–482.

Katon, W. J., & Walker, E. A. (1998). Medically unexplained symptoms in primary care. *Journal of Clinical Psychiatry, 59,* 15–21.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at eleven ages between 6.2 and 16.2 years. *Journal of Consulting and Clinical Psychology, 43,* 135–147.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Boston: Allyn & Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: Wiley.

Kaufman, A. S. (1999). *The effects of aging on WAIS-III IQs at ages 20 to 89 years.* Unpublished manuscript.

Kaufman, A. S., Kaufman, J. C., Balgopal, R., & McLean, J. E. (1996). Comparison of three WISC-III short forms: Weighing psychometric, clinical, and practical factors. *Journal of Clinical Child Psychology, 25,* 97–105.

Kaufman, A. S., & Kaufman, N. L. (1993). *Manual for the Kaufman Adolescent and Adult Intelligence Test (KAIT).* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (Eds.). (2002). *Specific learning disabilities: Psychological assessment and evaluation.* Cambridge, England: Cambridge University Press.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2002). *Assessing adolescent and adult intelligence.* Boston: Allyn & Bacon.

Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology, 44,* 231–248.

Kaufman, D. M. (2001). *Clinical neurology for psychiatrists* (5th ed.). New York: Saunders.

Kazdin, A. E. (1988). The diagnosis of childhood disorders: Assessment issues and strategies. *Behavioral Assessment, 10,* 67–94.

Keiller, S., & Graham, J. R. (1993). The meaning of low scores on the MMPI-2 clinical scales of normal subjects. *Journal of Personality Assessment, 61,* 211–223.

Keiser, R. E., & Prather, E. N. (1990). What is the TAT? A review of ten years of research. *Journal of Personality Assessments, 55,* 800–803.

Kelly, D., Marguilies, H., & Barrera, S. (1941). The stability of the Rorschach method as demonstrated in electroconvulsive therapy cases. *Rorschach Research Exchange, 5,* 44–48.

Kelly, E. L., & Fiske, D. W. (1951). *The prediction of performance in clinical psychology.* Ann Arbor: University of Michigan Press.

Keltikangas-Jarvinen, L. (1986). Concept of alexithymia. I: The prevalence of alexithymia in psychosomatic patients. *Psychotherapy and Psychosomatics, 44,* 132–138.

Kendall, P. C., & Hollon, S. D. (Eds.). (1981). *Assessment strategies for cognitive-behavioral interventions.* New York: Academic Press.

Kerns, L. L. (1986). Falsifications in the psychiatric history: A differential diagnosis. *Psychiatry, 49,* 13–17.

Killgore, W. D. S., & Dellapietra, L. (2000). Item response biases on the logical memory delayed recognition subtest of the Wechsler Memory Scale-III. *Psychological Reports, 86,* 851–857.

Kincel, R. L., & Murray, S. C. (1984). Kinesthesias in perception and the experience type: Dance and creative projection. *British Journal of Projective Psychology and Personality Study, 29,* 3–7.

Kinder, B. N. (1994). Where the action is in personality assessment. *Journal of Personality Assessment, 62,* 585–588.

Kipnis, D. (1968). Social immaturity, intellectual ability, and adjustive behavior in college. *Journal of Applied Psychology, 52,* 71–80.

Kirk, B. A., Cumming, R. W., & Hackett, H. H. (1963). Personal and vocational characteristics of dental students. *Personnel and Guidance Journal, 41,* 522–527.

Klein, R. G. (1986). Questioning the usefulness of projective psychological tests for children. *Journal of Developmental and Behavioral Pediatrics, 7,* 378–382.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107,* 296–310.

Klespies, P. M., & Dettmer, E. L. (2000). An evidence-based approach to evaluating and managing suicidal emergencies. *Journal of Clinical Psychology, 56,* 1109–1130.

Klinefelter, D., Pancoast, D. L., Archer, R. P., & Pruitt, D. L. (1990). Recent adolescent MMPI norms: T score elevation comparisons to Marks and Briggs. *Journal of Personality Assessment, 54,* 379–389.

Klinger, E. (1966). Fantasy need achievement as a motivational construct. *Psychological Bulletin, 66,* 291–308.

Klopfer, B. (1937). The present status of the theoretical development of the Rorschach method. *Rorschach Research Exchange, 1,* 142–147.

Klopfer, B., Ainsworth, M. D., Klopfer, W. G., & Holt, R. R. (1956). *Developments in the Rorschach technique* (Vol. 2). Yonkers, NY: World Books.

Klopfer, B., & Davidson, H. (1962). *The Rorschach technique: An introductory manual.* New York: Harcourt, Brace, Jovanovich.

Klopfer, W. G. (1960). *The psychological report.* New York: Grune & Stratton.

Klopfer, W. G. (1983). Writing psychological reports. In A. Walker (Ed.), *The handbook of clinical psychology: Theory, research, and practice.* Homewood, IL: Dow Jones-Irwin.

Knight, R. G., & Godfrey, H. P. D. (1996). Psychosocial aspects of neurological disorders: Implications for research in neuropsychology. *Australian Psychologist, 31,* 48–51.

Konold, T. R., Glutting, J. J., McDermott, P. A., Kush, J. C., & Watkins, M. W. (1999). Structure and diagnostic benefits of a normative taxonomy developed from the WISC-III standardization sample. *Journal of School Psychology, 37,* 29–48.

Koppitz, E. M. (1958). Relationships between the Bender-Gestalt Test and the Wechsler Intelligence Scale for Children. *Journal of Clinical Psychology, 14,* 413–416.

Koppitz, E. M. (1963). *The Bender Gestalt Test for Young Children* (Vol. 1). New York: Grune & Stratton.

Koppitz, E. M. (1975). *The Bender Gestalt Test for Young Children. Vol. 2: Research and Applications 1963–1973.* New York: Grune & Stratton.

Koss, M. P. (1979). MMPI item content: Recurring issues. In J. N. Butcher (Ed.), *New developments in the use of the MMPI* (pp. 3–38). Minneapolis: University of Minnesota Press.

Kostlan, A. (1954). A method for the empirical study of psychodiagnosis. *Journal of Consulting Psychology, 18,* 83–88.

Kramer, J. H. (1993). Interpretation of individual subtest scores on the WISC-III IQ and index scores. *Psychological Assessment, 5,* 193–196.

Kramer, J. H., Kaplan, E., & Huckeba, W. (1999). Configural errors on WISC-III Block Design. *Journal of the International Neuropsychological Society, 5,* 518–524.

Kratochwill, T. R. (1985). Selection of target behaviors in behavioral consultation. *Behavior Assessment, 7,* 49–61.

Krug, S. E. (1993). *Psychware sourcebook* (4th ed.). Kansas City, MO: Test Corporation of America.

Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., et al. (2000). Empirical support for psychological assessment in clinical care settings. *Professional Psychology, 31,* 119–130.

Kühn, R. (1963). Über die kritische Rorschach-Forschung and einige ihrer Ergebnisse. *Rorschachiana, 8,* 105–114.

Kunce, J. T., & Tamkin, A. S. (1981). Rorschach movement and color responses and MMPI social extraversion and thinking introversion personality types. *Journal of Personality Assessment, 45,* 5–10.

Kurtines, W., Hogan, R., & Weiss, D. (1975). Personality dynamics of heroin use. *Journal of Abnormal Psychology, 84,* 87–89.

Kurz, R. B. (1963). Relationship between time imagery and Rorschach human movement responses. *Journal of Consulting Psychology, 29,* 379–382.

Kvaal, S., Choca, J., & Groth-Marnat, G. (2003). The integrated psychological report. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrated assessment of adult personality* (2nd ed.). New York: Guilford Press.

Lacks, P. (1984). *Bender-Gestalt screening for brain dysfunction.* New York: Wiley.

Lacks, P. (1996). *Bender Gestalt Screening for Brain Dysfunction* [Computer software]. Odessa, FL: Psychological Assessment Resources.

Lacks, P. (1999). *Bender-Gestalt screening for brain dysfunction* (2nd ed.). New York: Wiley.

Lacks, P. (2000). Visuoconstructive abilities. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 401–436). New York: Wiley.

Lacks, P., & Newport, K. (1980). A comparison of scoring systems and level of scorer experience on the Bender-Gestalt Test. *Journal of Personality Assessment, 44,* 351–357.

Lafer, B. (1989). Predicting performance and persistance in hospice volunteers. *Psychological Reports, 65,* 467–472.

Lamb, D. G., Berry, D. T., Wetter, M. W., & Baer, R. A. (1994). Effects of two types of information on malingering of closed head injury on the MMPI-2: An analog investigation. *Psychological Assessment, 6,* 8–13.

Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G. M., et al. (1996). Administration and scoring manual for the Outcome Questionnaire (OQ45.2). Wilmington, DE: American Professional Credentialing Services.

Lanning, K. (1989). Detection of invalid response patterns on the California Psychological Inventory. *Applied Psychological Measurement, 13,* 45–56.

Lanyon, R. I. (1997). Detecting deception: Current models and directions. *Clinical Psychology: Science and Practice, 4,* 377–387.

Lanyon, R. I., & Goodstein, L. D. (1982). *Personality assessment* (2nd ed.). New York: Wiley.

Lapouse, R., & Monk, M. A. (1958). An epidemiologic study of behavior characteristics of children. *American Journal of Public Health, 48,* 1134–1144.

Lapouse, R., & Monk, M. A. (1964). Behavior deviations in a representative sample of children: Variations by sex, age, race, social class, and family size. *American Journal of Orthopsychiatry, 34,* 436–446.

Larrabee, G. J. (1986). Another look at VIQ-PIQ scores and unilateral brain damage. *International Journal of Neuroscience, 29,* 141–148.

La Rue, A., & Jarvik, L. R. (1987). Cognitive function and prediction of dementia in old age. *International Journal of Aging and Human Development, 25,* 78–89.

Laufer, W. S., Skoog, D. K., & Day, J. M. (1982). Personality and criminality: A review of the California Psychological Inventory. *Journal of Clinical Psychology, 38,* 562–573.

Lawrence, S. B. (1984). *Lawrence Psychological-Forensic Examination (Law-PSI).* San Bernadino, CA: Lawrence Psychological Center.

Lazarus, A. A. (1973). Multimodel behavior therapy: Treating the "BASIC ID." *Journal of Nervous and Mental Diseases, 156,* 404–411.

Lazarus, A. A. (1989). *The practice of multi-modal therapy.* Baltimore: Johns Hopkins University Press.

Leary, T. (1957). *Interpersonal diagnosis of personality.* New York: Ronald Press.

Leavitt, F., & Garron, G. C. (1982). Rorschach and pain characteristics of patients with low back pain and "conversion V" MMPI profiles. *Journal of Personality Assessment, 46,* 18–25.

Lefebvre, M. F. (1981). Cognitive distortions and cognitive errors in depressed and low back pain patients. *Journal of Consulting and Clinical Psychology, 49,* 517–525.

Lefkowitz, J., & Fraser, A. W. (1980). Assessment of achievement and power motivation of blacks and whites, using a black and white TAT with black and white administrators. *Journal of Applied Psychology, 65,* 685–696.

Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression: A meta-analytic approach. *Clinical Psychology Review, 21,* 401–419.

Leigh, J., & Zaylor, C. (2000). Cyberspace: Creating a therapeutic environment for telehealth applications. *Professional Psychology: Research and Practice, 31,* 478–483.

Lemsky, C. (2000). Neuropsychological assessment and treatment planning. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 535–576). New York: Wiley.

Leninger, B. E., Gramling, S. E., & Farrell, A. D. (1990). Neuropsychological deficits in symptomatic minor head injury patients after concussion and mild concussion. *Journal of Neurology, Neurosurgery, and Psychiatry, 53,* 293–296.

Lenzenweger, M. F. (1999). Stability and change in personality disorder features: The longitudinal study of personality disorders. *Archives of General Psychiatry, 56,* 1009–1015.

Leon, G. R., Gillum, B., Gillum, R., & Gouze, M. (1979). Personality stability and change over a 30-year period—middle age to old age. *Journal of Consulting and Clinical Psychology, 47,* 517–524.

Lerner, E. A. (1972). *The projective use of the Bender-Gestalt Test.* Springfield, IL: Charles C Thomas.

Lesiak, J. (1984). The Bender Visual Motor Gestalt Test: Implications for the diagnosis and prediction of reading achievement. *Journal of School Psychology, 22,* 391–405.

Leura, A. V., & Exner, J. E. (1978). *Structural differences in the records of adolescents as a function of being tested by one's own teacher* (Workshops Study No. 265, unpublished).

Levin, B. E., Llabre, M. M., & Weiner, W. J. (1988). Parkinson's disease and depression: Psychometric properties of the Beck Depression Inventory. *Journal of Neurology, Neurosurgery, and Psychiatry, 51,* 1401–1404.

Levin, H. S., High, W. M., Goethe, K. E., Sisson, R. A., Overall, J. E., Rhoades, H. M., et al. (1987). The Neurobehavioral Rating Scale: Assessment of the behavioral sequalae of head injury by the clinician. *Journal of Neurology, Neurosurgery, and Psychiatry, 50,* 183–193.

Levine, D. (1981). Why and when to test: The social context of psychological testing. In A. I. Rabin (Ed.), *Assessment with projective techniques* (pp. 553–580). New York: Springer.

Levitt, E. E., & Gotts, E. E. (1995). *The clinical application of MMPI special scales* (2nd ed.). Hillsdale, NJ: Erlbaum.

Levitt, E. E., & Truumaa, A. (1972). *The Rorschach technique with children and adolescents: Applications and norms.* New York: Grune & Stratton.

Lewak, R., & Hogan, L. (2003). Applying assessment information. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (2nd ed.). New York: Guilford Press.

Lewak, R. W., Marks, P. A., & Nelson, G. E. (1990). *Therapist guide to the MMPI and MMPI-2.* Muncie, IN: Accelerated Development.

Lewinsohn, P. M., Rohde, P., Klein, D. N., & Seeley, J. R. (1999). Natural course of adolescent major depressive disorder. I: Continuity into young adulthood. *Journal of the American Academy of Child Adolescent Psychiatry, 38,* 56–63.

Lezak, M. D. (1982). *The test-retest stability and reliability of some tests commonly used in neuropsychological assessment.* Paper presented at the 5th European conference of the International Neuropsychological Society, Deauville, France.

Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Lezak, M. D. (1988). IQ:RIP. *Journal of Experimental and Clinical Neuropsychology, 10,* 351–361.

Lezak, M. D. (1989a). *Assessment of the behavioral consequences of brain injury: Frontiers of clinical neuroscience* (Vol. 7). New York: Alan R. Liss.

Lezak, M. D. (1989b). Assessment of psychosocial dysfunctions resulting from head trauma. In M. D. Lezak (Ed.), *Assessment of the behavioral consequences of head trauma: Frontiers of clinical neuroscience* (Vol. 7). New York: Alan R. Liss.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lichtenberger, E. O., Kaufman, A. S., & Lai, Z. C. (2002). *Essentials of WMS-III assessment.* New York: Wiley.

Lichtenstein, S., & Fischoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20,* 159–183.

Light, B. H., & Amick, J. (1956). Rorschach responses of normal aged. *Journal of Projective Techniques, 20,* 185–195.

Lightfoot, S. L., & Oliver, J. M. (1985). The Beck Inventory: Psychometric properties in university students. *Journal of Personality Assessment, 49,* 434–436.

Lindgren, H. C., Moritsch, B., Thurlin, E. K., & Mich, G. (1986). Validity studies of three measures of achievement motivation. *Psychological Reports, 59,* 123–136.

Lindsay, D. S., & Read, J. D. (1995). "Memory work" and recovered memories of childhood sexual abuse: Scientific evidence and public and professional issues. *Psychology, Public Policy, and Law, 1,* 846–908.

Lindzey, G., & Herman, P. S. (1955). Thematic Apperception Test: A note on reliability and situational validity. *Journal of Projective Techniques, 19,* 36–42.

Lindzey, G., & Kalnins, D. (1958). Thematic Apperception Test: Some evidence bearing on the "hero assumption." *Journal of Abnormal and Social Psychology, 57,* 76–83.

Linscott, J., & DiGiuseppe, R. (1998). Cognitive assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 104–125). Boston: Allyn & Bacon.

Linton, H. B. (1954). Rorschach correlates of response to suggestion. *Journal of Abnormal and Social Psychology, 49,* 75–83.

Lipsitt, P. D., Lelos, D., & McGarry, A. L. (1971). Competency for trial: A screening instrument. *American Journal of Psychiatry, 128,* 137–141.

Little, K. B., & Shneidman, E. S. (1959). Congruencies among interpretations of psychological test and anamnestic data. *Psychological Monographs, 73* (6, Whole No. 476).

Little, S. G. (1992). The WISC-III: Everything old is new again. *School Psychology Quarterly, 7,* 138–142.

LoBello, S. G., Thompson, A. P., & Venugopala, E. (1998). Supplementary WAIS-III tables for determining subtest strengths and weaknesses. *Journal of Psychoeducational Assessment, 16,* 196–200.

Lodge, J., Tripp, G., & Harte, D. K. (2000). Think aloud, thought listing, and video-mediated recall procedures in the assessment of children's self talk. *Cognitive Therapy and Research, 24,* 399–418.

Loehelin, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist, 10,* 1285–1292.

Loenberger, L. T. (1989). The question of organicity: Is it still functional? *Professional Psychology: Research and Practice, 20,* 411–414.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

Lopez, E. C. (1997). The cognitive assessment of limited English proficient and bilingual children. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 503–518). New York: Guilford Press.

Loranger, A. W. (1988). *Personality Disorder Examination (PDE) manual.* Yonkers, NY: DV Communications.

Loranger, A. W., Sartorius, N., Andreoli, A., Berger, P., Buchheim, P., Channabasavanna, S. M., et al. (1995). The International Personality Disorder Examination. *Archives of General Psychiatry, 51,* 215–224.

Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1989). Verbal and visual memory index discrepancies from the Wechsler Memory Scale-Revised: Cautions in interpretation. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 198–202.

Loughmiller, G. C., Ellison, R. L., Taylor, C. W., & Price, P. B. (1970). Predicting career performances of physicians using the autobiographical inventory approach. *Proceedings of the American Psychological Association, 5,* 153–154.

Lovibund, P. F., & Lovibund, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scale (DASS) with the Beck Depression Inventories. *Behaviour Research and Therapy, 33,* 335–342.

Lowman, R. L. (1991). *The clinical practice of career assessment: Interests, abilities and personality.* Washington, DC: American Psychological Association.

Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935–1982. *American Psychologist, 39,* 451–454.

Lubin, B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1985). Psychological test usage patterns in five professional settings. *American Psychologist, 40,* 857–861.

Lubin, B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1986). Selected characteristics of psychologists and psychological assessment in five settings: 1959–1988. *Professional Psychology: Research and Practice, 17,* 155–157.

Luborsky, L. (1994). Therapeutic alliances as predictors of psychotherapy outcomes: Factors explaining predictive success. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 38–50). New York: Wiley.

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies. *Archives of General Psychiatry, 32,* 995–1008.

Lucas, C. P., Zhang, H., Fisher, P. W., Shaffer, D., Regier, D. A., Narow, W. E., et al. (2001). The DISC Predictive Scales (DPS): Efficiently screening for diagnosis. *Journal of the Academy of Child and Adolescent Psychiatry, 40,* 443–449.

Luckasson, R., Coulter, D. L., Polloway, E. A., Reiss, S., Schalock, R. L., Snell, M. E., et al. (1992). *Mental retardation: Definition, classification, and systems of support.* Washington, DC: American Association on Mental Retardation.

Lundy, A. (1985). The reliability of the Thematic Apperception Test. *Journal of Personality Assessment, 49,* 141–145.

Lundy, A. (1988). Instructional set and Thematic Apperception Test validity. *Journal of Personality Assessment, 52,* 309–320.

Luria, A. R. (1973). *The working brain.* New York: Basic Books.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York: Basic Books.

MaCartney-Filgate, M. S., & Vriezen, E. R. (1988). Intercorrelation of clinical tests of verbal memory. *Archives of Clinical Neuropsychology, 3,* 121–126.

Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "intelligent testing" approach to the WISC-III. *School Psychology Quarterly, 12,* 197–234.

Maddox, T. (1997). *Tests: A comprehensive reference for assessments in psychology, education, and business* (4th ed.). Austin, TX: ProEd.

Maitra, A. K. (1983). Executive effectiveness: Characteristic thematic phantasy. *Managerial Psychology, 4,* 59–68.

Majumber, A. K., & Roy, A. B. (1962). Latent personality content of juvenile delinquents. *Journal of Psychological Research, 1,* 4–8.

Makatura, T. J., Lam, C. S., Leahy, B. J., Castillo, M. T., & Kalpakjian, C. Z. (1999). Standardized memory tests and the appraisal of everyday memory. *Brain Injury, 13,* 355–367.

Mallinckrodt, B. (1996). Change in working alliance, social support, and psychological symptoms in brief therapy. *Journal of Counseling Psychology, 43,* 448–455.

Maloney, M. P., & Ward, M. P. (1976). *Psychological assessment: A conceptual approach*. New York: Oxford University Press.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13,* 330–334.

Mann, L. (1956). The relation of Rorschach indices of extratension and introversion to a measure of responsiveness to the immediate environment. *Journal of Consulting Psychology, 20,* 114–118.

Marchese, M. C., & Muchinsky, P. M. (1993). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment, 1,* 18–26.

Margolin, D. I., Pate, D. S., Friedrich, F. J., & Elia, E. (1990). Dysnomia in dementia and in stroke patients: Different underlying cognitive deficits. *Journal of Clinical and Experimental Neuropsychology, 12,* 597–612.

Margolin, G., Hattem, D., John, R. S., & Yost, K. (1985). Perceptual agreement between spouses and outside observers when coding themselves and a stranger dyad. *Behavioral Assessment, 7,* 235–247.

Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The actuarial use of the MMPI with adolescents and adults.* Baltimore: Williams & Wilkins.

Marley, M. L. (1982). *Organic brain pathology and the Bender Gestalt Test: A differential diagnostic scoring system.* New York: Grune & Stratton.

Marlowe, D. B., Festinger, D. S., & Kirby, K. C. (1998). Congruence of the MCMI-II and MCMI-III in cocaine dependence. *Journal of Personality Assessment, 71,* 15–28.

Martin, J. D., Pfaadt, N. K., & MaKinster, J. G. (1983). Relationship of hostility and white space responses on the Rorschach. *Perceptual and Motor Skills, 57,* 739–742.

Martin, P., Churchard, M., Kotcher, S., & Korenblum, M. (1991). Diagnostic utility of the Beck Depression Inventory with adolescent psychiatric outpatients and inpatients. *Canadian Journal of Psychiatry, 36,* 428–431.

Maruish, M. E. (Ed.). (1999). *Use of psychological testing for treatment planning and outcome assessment* (2nd ed.). Mahwah, NJ: Erlbaum.

Maruish, M. E. (Ed.). (2000). *Handbook of psychological testing in primary care settings.* Mahwah, NJ: Erlbaum.

Maruish, M. E. (2002). *Psychological testing in the age of managed behavioral health care.* Mahwah, NJ: Erlbaum.

Masling, J. (1992). The influence of situation and interpersonal variables in projective testing. *Journal of Personality Assessment, 59,* 616–640.

Mason, B., & Exner, J. E. (1984). *Correlations between WAIS subtests and nonpatient adult Rorschach data.* Rorschach Workshops (Study No. 289, unpublished).

Massman, P. J., & Bigler, E. D. (1993). A quantitative review of the diagnostic utitiity of the WAIS-R Fuld Profile. *Archives of Clinical Neuropsychology, 8,* 417–428.

Massman, P. J., Delis, D. C., Butters, N., Dupont, M., & Gillin, N. (1992). The subcortical dysfunction model of memory deficits in depression: Neuropsychological validation in a subgroup of patients. *Journal of Clinical and Experimental Neuropsychology, 14,* 687–706.

Matarazzo, J. D. (1965). The interview. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 403–450). New York: McGraw-Hill.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist, 45,* 999–1017.

Matarazzo, J. D., Daniel, M. H., Prifitera, A., & Herman, D. O. (1988). Inter-subtest scatter in the WAIS-R standardization sample. *Journal of Clinical Psychology, 44,* 940–950.

Matarazzo, J. D., & Prifitera, A. (1989). Subtest scatter and premorbid intelligence: Lessons from the WAIS-R standardization sample. *Psychological Assessment, 1,* 186–191.

Mauger, P. A. (1972). The test-retest reliability of persons: An empirical investigation utilizing the MMPI and the personality research form. *Dissertation Abstracts International, 33,* 2816B.

Mayes, S. D., Calhoun, S. L., & Crowell, E. W. (1998). WISC-III profiles for children with and without learning disabilities. *Psychology in the Schools, 35,* 309–316.

Maziade, M., Roy, A. A., Fournier, J. P., Cliche, D., Merette, C., Caron, C., et al. (1992). Reliability of best-estimate diagnosis in genetic linkage studies of major psychoses. *American Journal of Psychiatry, 149,* 1674–1686.

McAllister, L. (1996). *A practical guide to CPI interpretation* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

McCall, R. B., Appelbaum, M. I., & Hogarty, P. S. (1973). Developmental changes in mental performance. *Monographs of the Society for Research in Child Development, 38*(3, Serial No. 150), 1–83.

McCallum, R. S., & Bracken, B. A. (1997). The Universal Nonverbal Intelligence Test. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 268–280). New York: Guilford Press.

McCallum, R. S., Bracken, B. A., & Wasserman, J. (2001). *Essentials of nonverbal assessment.* New York: Wiley.

McCann, J. T. (1991). Convergent and discriminant validity of the MCMI-II and MMPI personality disorder scales. *Psychological Assessment, 3,* 9–18.

McClelland, D. C. (1961). *The achieving society.* Princeton, NJ: Van Nostrand.

McClelland, D. C. (1966). Longitudinal trends in the relation of thought to action. *Journal of Consulting Psychology, 30,* 479–483.

McClelland, D. C. (1971). *Assessing human motivation.* New York: General Learning Press.

McClelland, D. C., Koestner, P., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96,* 690–702.

McConnaughy, E. A., Prochaska, J. O., & Velicer, W. (1983). Stages of change in psychotherapy: Measurement and sample profiles. *Psychotherapy: Theory, Research, and Practice, 20,* 368–375.

McConnell, O. L. (1967). Koppitz's Bender Gestalt in relation to organic and emotional problems in children. *Journal of Clinical Psychology, 23,* 370–374.

McCormick, I. A. (1984). A simple version of the Rathus Assertiveness Schedule. *Behavioral Assessment, 7,* 95–99.

McCown, W., Fink, A. D., Galina, H., & Johnson, J. (1992). Effects of laboratory-induced controllable and uncontrollable stress on Rorschach variables m and Y. *Journal of Personality Assessment, 59,* 564–573.

McCracken, L. M., & Franzen, M. D. (1992). Principal-components analysis of the equivalence of alternate forms of the Trail Making Test. *Psychological Assessment, 4,* 235–238.

McCrae, R. R., Costa, P. T., & Piedmont, R. L. (1993). Folk concepts, natural language, and psychological constructs: The California Psychological Inventory and the five-factor model. *Journal of Personality Assessment, 61,* 1–25.

McCurry, S. M., Fitz, A. G., & Terri, L. (1994). Comparison of age-extended norms for the Wechsler Adult Intelligence Scale-Revised in patients with Alzheimer's disease. *Psychological Assessment, 6,* 231–235.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599–616.

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, M. (1992). Illusions of meaning in the ipsative assessment of children's abilities. *Journal of Special Education, 25,* 504–526.

McFall, R. M., & Lillesand, D. V. (1971). Behavior rehearsal with modeling and coaching in assertive training. *Journal of Abnormal Psychology, 77,* 313–323.

McFarland, R. A. (1984). Effects of music upon emotional content of TAT stories. *Journal of Psychology, 116,* 227–234.

McFie, J. (1960). Psychological testing in clinical neurology. *Journal of Nervous and Mental Diseases, 131,* 383–393.

McFie, J. (1969). The diagnostic significance of disorders of higher nervous activity syndromes related to frontal, temporal, parietal, and occipital lesions. In P. J. Vinken & G. W. Bruyn (Eds.), *Handbook of clinical neurology* (Vol. 4). New York: American Elsevier.

McGlyn, F. D., & Rose, M. P. (1998). Assessment of anxiety and fear. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 179–209). Boston: Allyn & Bacon.

McGrath, R. E. (2001). Toward more clinically relevant research. *Journal of Personality Assessment, 77,* 307–332.

McGrath, R. E., & Ingersoll, J. (1999a). Writing a good cookbook. I: A review of MMPI high-point code system studies. *Journal of Personality Assessment, 73,* 149–178.

McGrath, R. E., & Ingersoll, J. (1999b). Writing a good cookbook. II: A synthesis of MMPI high-point code system study effect sizes. *Journal of Personality Assessment, 73,* 179–198.

McGrath, R. E., Powis, D., & Pogge, D. L. (1998). Code type-specific tables for interpretation of MMPI-2 Harris and Lingoes Subscales: Consideration of gender and code definition. *Journal of Clinical Psychology, 54,* 655–664.

McIntosh, J. A., Belter, R. W., Saylor, C. F., Finch, A. J., & Edwards, G. L. (1988). The Bender-Gestalt with adolescents: Comparison of two scoring systems. *Journal of Clinical Psychology, 44,* 226–230.

McLean, P., & Taylor, S. (1992). Severity of unipolar depression and choice of treatment. *Behavior Research and Therapy, 30,* 443–451.

McLeod, C. C., Budd, M. A., & McClelland, D. C. (1997). Treatment of somatization in primary care. *General Hospital Psychiatry, 19,* 251–258.

McMinn, M. R., Buchanan, T., Ellens, B. M., & Ryan, M. K. (1999). Technology, professional practice, and ethics: Survey findings and implications. *Professional Psychology: Research and Practice, 30,* 165–172.

McMinn, M. R., Ellens, B. M., & Soref, E. (1999). Ethical perspectives and practice behaviors involving computer-based test interpretation. *Assessment, 6,* 71–77.

McNulty, J. L., Ben-Porath, Y. S., & Graham, J. R. (1998). An empirical examination of the correlates of well-defined and not-defined MMPI-2 code types. *Journal of Personality Assessment, 71,* 393–410.

McNulty, J. L., Graham, J. R., Ben-Porath, Y. S., & Stein, L. A. R. (1997). Comparative validity of MMPI-2 scores of African American and Caucasian mental health center clients. *Psychological Assessment, 9,* 464–470.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1965). Seer over sign: The first good example. *Journal of Experimental Research in Personality, 1,* 27–32.

Megargee, E. I. (1964). Undercontrol and overcontrol in assaultive and homicidal adolescents (Doctoral dissertation, University of California, Berkley, 1964). *Dissertation Abstracts International* (UMI No. 64–9923).

Megargee, E. I. (1965). Assault with intent to kill. *Trans-Action, 2,* 27–31.

Megargee, E. I. (1966). Undercontrolled and overcontrolled personality types in extreme anti-social aggression. *Psychological Monographs, 80*(611).

Megargee, E. I. (1972). *The California Psychological Inventory handbook.* San Francisco: Jossey-Bass.

Megargee, E. I. (1995). Assessing and understanding aggressive and violent patients. In J. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 395–409). New York: Oxford University Press.

Megargee, E. I. (2002). *California Psychological Inventory handbook* (2nd ed.). San Francisco: Jossey-Bass.

Megargee, E. I., Cook, P. E., & Mendelsohn, G. A. (1967). Development and evaluation of an MMPI scale of assaultiveness in overcontrolled individuals. *Journal of Abnormal Psychology, 72,* 519–528.

Megargee, E. I., & Mendelsohn, G. A. (1962). A cross-validation of twelve MMPI indices of hostility and control. *Journal of Abnormal Psychology, 65,* 431–438.

Megargee, E. I., & Parker, G. V. (1968). An exploration of the equivalence of Murrayan needs as assessed by the Adjective Check List, the TAT, and the Edwards Personal Preference Schedule. *Journal of Clinical Psychology, 24,* 47–51.

Mehrabian, A. (1972). *Nonverbal communication.* Chicago: Aldine-Atherton.

Melzack, R. (1975). The McGill Pain Questionnaire: Major properties and scoring methods. *Pain, 1,* 277–299.

Mendez, F. (1978). *Adult Neuropsychological Questionnaire.* Odessa, FL: Psychological Assessment Resources.

Mendoza, J. (2001). Reporting the results of the neuropsychological evaluation. In C. G. Armengol, E. Kaplan, & E. J. Moes (Eds.), *The consumer-oriented neuropsychological report* (pp. 95–122). Lutz, FL: Psychological Assessment Resources.

Mercer, J. R. (1979). In defense of racially and culturally non-discriminatory assessment. *School Psychology Digest, 8,* 89–115.

Mercer, J. R., & Lewis, J. F. (1978). *System of multicultural pluralistic assessment.* San Antonio, TX: Psychological Corporation.

Mermelstein, J. J. (1983). The relationship between rotation on the Bender-Gestalt Test and ratings of patient disorientation. *Journal of Personality Assessment, 47,* 490–491.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Psychological Assessment, 7,* 741–749.

Meyer, G. J. (1997). On the integration of personality assessment methods: The Rorschach and the MMPI. *Journal of Personality Assessment, 68,* 297–330.

Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Research, 75,* 46–81.

Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment, 13,* 486–502.

Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment, 69,* 1–38.

Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Fowler, J. C., Piers, C. C., et al. (2002). An examination of interrater reliability for scoring the Rorschach Comprehensive System in eight data sets. *Journal of Personality Assessment, 78,* 219–274.

Meyer, P., & Davis, S. (1992). *The CPI applications guide: An essential tool for individual, group, and organizational development.* Palo Alto, CA: Consulting Psychologists Press.

Meyer, R. G. (1996). *The clinician's handbook: Integrated diagnostics, assessment and intervention in adult and adolescent psychopathology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Meyerink, L. H., Reitan, R. M., & Selz, M. (1988). The validity of the MMPI with multiple sclerosis. *Journal of Clinical Psychology, 44,* 764–769.

Meyers, J. E., & Meyers, K. R. (1996). *Rey Complex Figure Test and recognition trial: A professional manual.* Odessa, FL: Psychological Assessment Resources.

Miceli, G., Caltagirone, C., Gainotti, G., Masullo, C., & Silveri, M. C. (1981). Neuropsychological correlates of localized cerebral lesions in non-aphasic brain-damaged patients. *Journal of Clinical Neuropsychology, 3,* 53–63.

Milberg, W. P., Hebben, N., & Kaplan, E. (1996). The Boston Process Approach to neuropsychological assessment. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (2nd ed., pp. 58–80). New York: Oxford University Press.

Miller, H. R., Streiner, D. L., & Parkinson, A. (1992). Maximum likelihood estimates of the ability of the MMPI and the MCMI personality disorder scales and the SIDP to identify personality disorders. *Journal of Personality Assessment, 59,* 1–13.

Miller, L. (1993, January/February). Toxic torts: Clinical neuropsychological and forensic aspects of chemical and electrical injuries. *Journal of Cognitive Rehabilitation, 11,* 6–18.

Millis, S. R., Malina, A. C., Bowers, D. A., & Ricker, J. H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale-III. *Journal of Clinical and Experimental Neuropsychology, 21,* 87–93.

Millon, T. (1969). *Modern psychopathology: A biosocial approach to maladaptive learning and functioning.* Philadelphia: Saunders.

Millon, T. (1977). *Millon Clinical Multiaxial Inventory.* Minneapolis, MN: National Computer Systems.

Millon, T. (1987). *Manual for the MCMI-II* (2nd ed.). Minneapolis, MN: National Computer Systems.

Millon, T. (1992). Millon Clinical Multiaxial Inventory: I & II. *Journal of Counseling and Development, 70,* 422–426.

Millon, T. (1993). *The MACI manual.* Minneapolis, MN: National Computer Systems.

Millon, T. (1994). *Manual for the MCMI-III.* Minneapolis, MN: National Computer Systems.

Millon, T. (1997). *Millon Clinical Multiaxial Inventory-III manual* (2nd ed.). Minneapolis, MN: National Computer Systems.

Millon, T., & Davis, R. D. (1996). *Disorders of personality: DSM-IV and beyond* (2nd ed.). New York: Wiley.

Millon, T., Green, C. J., & Meagher, R. B. (2000). *Manual for the Millon Behavioral Health Inventory.* Minneapolis, MN: National Computer Systems.

Mills, C. J., & Bohannon, W. E. (1980). Personality characteristics of effective state police officers. *Journal of Applied Psychology, 65,* 680–684.

Milne, D. (1984). Improving the social validity and implementation of behavior therapy training for psychiatric nurses using a patient-centered learning format. *British Journal of Clinical Psychology, 23,* 313–314.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Mitrushina, M., Boone, K., & D'Elia, L. F. (1999). *Handbook of normative data for neuropsychological assessment.* New York: Oxford University Press.

Mitrushina, M., Satz, P., & Van Gorp, W. (1989). Some putative cognitive precursors in subjects hypothesized to be at-risk for dementia. *Archives of Clinical Neuropsychology, 4,* 323–333.

Mittenberg, W., Hammeke, T. A., & Rao, S. M. (1989). Intrasubtest scatter on the WAIS-R as a pathognomonic sign of brain injury. *Psychological Assessment, 1,* 273–276.

Mizes, J. S., & Christiano, B. A. (1994). Assessment of cognitive variables relevant to cognitive behavioral perspectives on anorexia nervosa and bulimia nervosa. *Behavior Research and Therapy, 33,* 95–105.

Mizushima, K., & DeVos, G. (1967). An application of the California Psychological Inventory in a study of Japanese delinquency. *Journal of Clinical Psychology, 71,* 45–51.

Mohr, D. C. (1995). The role of proscription in psychotherapy. *Psychotherapy, 32,* 187–193.

Molish, H. B. (1967). Critique and problems of the Rorschach: A survey. In S. J. Beck & H. B. Molish (Eds.), *Rorschach's test. Vol. 2: A variety of personality pictures* (2nd ed., pp. 45–48). New York: Grune & Stratton.

Monahan, J., & Steadman, H. J. (2001). Violence risk assessment: A quarter century of research. In L. E. Frost & R. J. Bonnie (Eds.), *The evolution of mental health law* (pp. 195–211). Washington, DC: American Psychological Association.

Monahan, J., Steadman, H. J., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., Silver, E., et al. (2000). Developing a clinically useful actuarial tool for assessing violence risk. *British Journal of Psychiatry, 176,* 312–319.

Moon, G. W., Blakey, W. A., Gorsuch, R. L., & Fantuzzo, J. W. (1991). Frequent WAIS-R administration errors: An ignored source of inaccurate measurement. *Professional Psychology, 22,* 256–258.

Moon, G. W., Fantuzzo, J. W., & Gorsuch, R. L. (1986). Teaching WAIS-R administration skills: Comparison of the MASTERY model to other existing clinical training modalities. *Professional Psychology, 17,* 31–35.

Moos, R. H. (1990). Depressed patients' life context, amount of treatment, and treatment outcome. *Journal of Nervous and Mental Diseases, 178,* 105–112.

Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications: A research-based measurement procedure. *American Psychologist, 50,* 14–23.

Morena, D. (1981). The healthy drawing. In G. Groth-Marnat & D. Morena (Eds.), *Handbook of psychological assessment* (pp. 80–85). Unpublished manuscript.

Morey, L. C., Blashfield, R. K., Webb, W. W., & Jewell, J. (1988). MMPI scales for *DSM-III* personality disorders: A preliminary validation study. *Journal of Clinical Psychology, 44,* 47–50.

Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies. *AMA Archives of Neurology and Psychiatry, 34,* 389–406.

Morgan, G., Gliner, J. A., & Harmon, R. J. (2001). Measurement validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 729–731.

Morlan, K. K., & Tan, S.-Y. (1998). Comparison of the Brief Psychiatric Rating Scale and the Brief Symptom Inventory. *Journal of Clinical Psychology, 54,* 885–894.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., et al. (1989). The consortium to establish a registry for Alzheimer's disease

(CERAD). Part 1: Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39,* 1159–1165.

Morrow, L. A., Furman, J. M. R., Ryan, C. M., & Hodgson, M. J. (1988). Neuropsychological deficits associated with verbatim abnormalities in solvent exposed workers. *Clinical Neuropsychologist, 2,* 272–273.

Mostofsky, D., & Barlow, D. H. (2000). *The management of stress and anxiety in medical disorders.* Needham Heights, MA: Allyn & Bacon.

Mueller, H. H., Dash, V. N., Matheson, D. W., & Short, R. H. (1984). WISC-R subtest patterning of below average, average, and above average I.Q. children: A meta-analysis. *Alberta Journal of Educational Research, 30,* 68–85.

Mukerji, M. (1969). Rorschach indices of love, aggression, and happiness. *Journal of Projective Techniques and Personality Assessment, 33,* 526–529.

Mulvey, E., & Cauffman, E. (2001). The inherent limits of predicting school violence. *American Psychologist, 56,* 797–802.

Munley, P. H. (1991). Confidence intervals for the MMPI-2. *Journal of Personality Assessment, 57,* 52–60.

Murdoch, B. E., Chenery, H. J., Wilks, V., & Boyle, R. S. (1987). Language disorder in dementia of the Alzheimer's type. *Brain and Language, 31,* 122–137.

Murillo, L. G., & Exner, J. E. (1973). The effects of regressive ECT with process schizophrenics. *American Journal of Psychiatry, 130,* 269–273.

Murphy, J. M., Monson, R. R., Laird, N. M., Sobol, A. M., & Leighton, A. H. (2000). A comparison of diagnostic interviews for depression in the Stirling County Study: Challenges for psychiatric epidemiology. *Archives of General Psychiatry, 57,* 230–236.

Murphy, L. L., Impara, J. C., & Plake, B. S. (Eds.). (1999). *Tests in print V.* Lincoln, NE: Buros Institute of Mental Measurements.

Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.

Murray, H. A. (1943). *Thematic Apperception Test manual.* Cambridge, MA: Harvard University Press.

Murstein, B. I. (1960). Factor analysis of the Rorschach. *Journal of Consulting Psychology, 24,* 262–275.

Murstein, B. I. (1963). *Theory and research in projective techniques (Emphasizing the TAT).* New York: Wiley.

Naglieri, J. A. (1993). Pairwise and ipsative comparisons of WISC-III IQ and index scores. *Psychological Assessment, 5,* 113–116.

Naglieri, J. A. (1999). *Essentials of CAS assessment.* New York: Wiley.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System Administration and Scoring manual.* Itasca, IL: Riverside.

Nathan, P. E., & Gorman, J. M. (Eds.). (1998). *A guide to treatments that work.* New York: Oxford University Press.

Neale, M. D., & McKay, M. F. (1985). Scoring the Bender-Gestalt Test using the Koppitz Developmental System: Interrater reliability, item difficulty, and scoring implications. *Perceptual and Motor Skills, 60,* 627–636.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Nelson, H. E., & Williams, J. R. (1991). *National Adult Reading Test (NART): Test manual.* Windsor, England: NFER-Nelson.

Nelson, J. R., Roberts, M. L., Rutherford, R. B., Mathur, S. R., & Aaroe, L. A. (1999). A statewide survey of special education administrators and school psychologists regarding functional behavioral assessment. *Education and Treatment of Children, 22,* 267–279.

Nelson, L. D., Satz, P., Mitrushina, M., Van Gorp, W., Cicchetti, D., Lewis, R., et al. (1989). Development and validation of the Neuropsychology Behavior and Affect Profile. *Psychological Assessment, 1,* 266–272.

Nelson, R. E., & Maser, J. D. (1988). The *DSM-III* and depression: Potential contributions of behavioral assessment. *Behavioral Assessment, 10,* 45–66.

Nelson-Gray, R. O., & Farmer, R. F. (1999). Behavioral assessment of personality disorders. *Behavior Research and Therapy, 37,* 347–368.

Newcomb, F. (1969). *Missile wounds of the brain.* London: Oxford University Press.

Newman, F. L., Ciarlo, J. A., & Carpenter, D. (1999). Guidelines for selecting psychological instruments for treatment planning and outcome assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 153–170). Mahwah, NJ: Erlbaum.

Newman, J. L., Gray, E. A., & Fuqua, D. R. (1998). The relation between career indecision to personality dimension of the California Psychological Inventory. *Journal of Vocational Behavior, 54,* 174–187.

Nicholi, A. M. (1978). History and mental status. In A. M. Nicholi (Ed.), *The Harvard guide to modern psychiatry.* Cambridge, MA: Harvard University Press.

Nilson, L., Barregard, L., & Baeckman, L. (1999). Trail Making Test in chronic toxic encephalopath: Performance and discriminative potential. *Clinical Neuropsychologist, 13,* 314–327.

Norcross, J. C. (1997). Emerging breakthroughs in psychotherapy integration: Three predictions and one fantasy. *Psychotherapy, 34,* 86–90.

Norcross, J. C. (Ed.). (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient's needs.* New York: Oxford University Press.

Norcross, J. C., & Beutler, L. E. (1997). Determining the therapeutic relationship of choice in brief therapy. In J. N. Butcher (Ed.), *Objective psychological assessment in managed health care: A practitioner's guide* (pp. 42–60). New York: Oxford University Press.

Norton, J. C. (1978). The Trail Making Test and Bender Background Interference Procedure as screening devices. *Journal of Clinical Psychology, 34,* 916–922.

Nottingham, E. J., & Mattson, R. E. (1981). A validation study of the Competency Screening Test. *Law and Human Behavior, 5,* 329–335.

Oakland, T. D. (1980). An evaluation of the ABIC, pluralistic norms, and estimated learning potential. *Journal of School Psychology, 18,* 3–11.

Oas, P. (1984). Validity of the Draw-A-Person and Bender Gestalt as measures of impulsivity with adolescents. *Journal of Counsulting and Clinical Psychology, 52,* 1011–1019.

Ober, B. A., Koss, E., Friedland, R. P., & Delis, D. C. (1985). Processes of verbal memory failure in Alzheimer-type dementia. *Brain and Cognition, 4,* 90–103.

Obrzut, J. E., & Cummings, J. A. (1983). The projective approach to personality assessment: An analysis of thematic picture techniques. *School Psychology Review, 12,* 414–420.

Office of Science and Technology. (1967). *Privacy and behavioral research.* Washington, DC: U.S. Government Printing Office.

Office of Strategic Services Staff. (1948). *Assessment of men.* New York: Holt, Rinehart and Winston.

Okun, A., Stein, R. K., Bauman, L. J., & Silver, E. J. (1996). Content validity of the Psychiatric Symptom Index, CES Depression Scale, and State-Trait Anxiety Inventory from the perspective of *DSM-IV. Psychological Reports, 79,* 1059–1069.

O'Leary, A. (1985). Self-efficacy and health. *Behavior Therapy and Research, 23,* 437–451.

Olin, J. T., Schneider, L. S., Eaton, E., Zemansky, M. F., & Pollock, V. E. (1992). The Geriatric Depression Scale and the Beck Depression Inventory as screening instruments in an older adult outpatient population. *Psychological Assessment, 4,* 190–192.

Ollendick, T. H. (1978). *The Fear Survey Schedule for Children-Revised.* Unpublished manuscript, Indiana State University, Terre Haute.

Ollendick, T. H. (1983). Reliability and validity of the revised Fear Survey Schedule for Children (FSSC-R). *Behavior Research and Therapy, 21,* 685–692.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 766–783.

Osman, A., Downs, W., Barrios, F. X., Kopper, B. A., Gutierrez, P. M., & Chiros, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory-III. *Journal of Psychopathology and Behavioral Assessment, 19,* 359–376.

Osterrith, P. A. (1944). Le test de copie d'une figure complexe. *Archives de Psychologie, 30,* 206–356.

Othmer, E., & Othmer, S. C. (1994). *The clinical interview using DSM-IV. Vol. 1: Fundamentals.* Washington, DC: American Psychiatric Press.

Otto, R. K., & Heilburn, K. (2002). The practice of forensic psychology: A look toward the future in light of the past. *American Psychologist, 57,* 5–10.

Ownby, R. L. (1990). A study of the expository process model in mental health settings. *Journal of Clinical Psychology, 46,* 366–371.

Ownby, R. L. (1997). *Psychological reports: A guide to report writing in professional psychology* (3rd ed.). Brandon, VT: Clinical Psychology.

Ownby, R. L., & Wallbrown, F. H. (1983). Evaluating school psychological reports. Part I: A procedure for systematic feedback. *Psychology in the Schools, 20,* 41–45.

Page, H. A. (1957). Studies in fantasy-daydreaming frequency and Rorschach scoring categories. *Journal of Consulting Psychology, 21,* 111–114.

Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1998). Base rates of "impaired" neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology, 13,* 503–511.

Palmer, J. O. (1970). *The psychological assessment of children.* New York: Wiley.

Pancoast, D. L., & Archer, R. P. (1988). MMPI adolescent norms: Patterns and trends across four decades. *Journal of Personality Assessment, 52,* 691–706.

Pantle, M. L., Ebner, D. L., & Hynan, L. S. (1994). The Rorschach and the assessment of impulsivity. *Journal of Clinical Psychology, 50,* 633–638.

Parker, K. C. H. (1983). A meta-analysis of the reliability and validity of the Rorschach. *Journal of Personality Assessment, 47,* 227–231.

Parker, K. C. H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin, 103,* 367–373.

Parks, C. W. (1982). *A multi-dimensional view of the imagery construct: Issues of definition and assessment.* Unpublished manuscript.

Parks, C. W., & Hollon, S. D. (1988). Cognitive assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3rd ed., pp. 161–212). New York: Pergamon Press.

Parks, R. W., Loewenstein, D. A., Dodrill, K. L., Barker, W. W., Yoshii, F., Chang, J. Y., et al. (1988). Cerebral metabolic effects of a verbal fluency test: A PET scan study. *Journal of Clinical and Experimental Neuropsychology, 10,* 565–575.

Pascal, G. R., & Suttell, B. J. (1951). *The Bender Gestalt Test: Quantification and validity for adults.* New York: Grune & Stratton.

Patterson, C. H. (1989). Foundations for a systematic eclectic psychotherapy. *Psychotherapy, 26,* 427–435.

Paul, G. L. (1967). Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Handbook of psychotherapy integration* (pp. 300–334). New York: Basic Books.

Paunonen, S. V., & Ashton, M. C. (1998). The structured assessment of personality across cultures. *Journal of Cross-Cultural Psychology, 29,* 150–170.

Pekarik, G., & Wolff, C. B. (1996). Relationship of satisfaction to symptom change, follow-up adjustment, and clinical significance. *Professional Psychology: Research and Practice, 27,* 202–208.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77,* 1296–1312.

Pennebaker, J. W., & Seagal, J. D. (1999). Forming a story: The health benefits of narrative. *Journal of Clinical Psychology, 55,* 1243–1254.

Perez, R. G., Ascaso, E. E., Massons, J. M. D., & Chaparro, N. (1998). Characteristics of the subject and interview influencing the test-retest reliability of the Diagnostic Interview for Children and Adolescents-Revised. *Journal of Child Psychology and Psychiatry, 39,* 963–972.

Perkins, D. N., & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist, 52,* 1125–1133.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behavior. *Neuropsychologia, 12,* 323–330.

Perry, J. C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry, 149,* 1645–1653.

Perry, W., Sprock, J., Schaible, D., McDougall, A., Minassian, A., Jenkins, M., et al. (1995). Amphetamine use on Rorschach measures in normal subjects. *Journal of Personality Assessment, 64,* 456–465.

Persons, J. B., Mooney, K., & Padesky, C. A. (1995). Interrater reliability of cognitive behavioral case formulations. *Cognitive Therapy and Research, 19,* 21–34.

Petrocelli, J. V. (2002). Processes and stages of change: Counseling with the transtheoretical model of change. *Journal of Counseling and Development, 80,* 22–30.

Peveler, R. C., & Fairburn, C. G. (1990). Measurement of neurotic symptoms by self report questionnaire: Validity of the SCL-90-R. *Psychological Medicine, 20,* 873–879.

Pfeifer, C., & Sedlacek, W. (1971). The validity of academic predictors for black and white students at a predominantly white university. *Journal of Educational Measurement, 8,* 253–261.

Phillips, L., & Smith, J. (1953). *Rorschach interpretation: Advanced technique.* New York: Grune & Stratton.

Piedmont, R. L., Sokolov, R. L., & Flemming, M. Z. (1989a). On WAIS-R Difference Scores in a psychiatric population. *Psychological Assessment, 1,* 155–159.

Piedmont, R. L., Sokolov, R. L., & Flemming, M. Z. (1989b). An examination of some diagnostic strategies involving the Wechsler intelligence scales. *Psychological Assessment, 1,* 181–185.

Piersma, H. L. (1986). The stability of the Millon Clinical Multiaxial Inventory for psychiatric patients. *Journal of Personality Assessment, 50,* 193–197.

Piersma, H. L., Boes, J. L., & Reaume, W. M. (1994). Unidimensionality of the Brief Symptom Inventory (BSI) in adult and adolescent inpatients. *Journal of Personality Assessment, 63,* 338–344.

Pilowski, I., Spence, N., Cobb, J., & Katsikitis, M. (1984). The Illness Behavior Questionnaire as an aid in clinical assessment. *General Hospital Psychiatry, 6,* 123–130.

Piotrowski, C. (1984). The status of projective techniques: Or, "Wishing won't make it go away." *Journal of Clinical Psychology, 40,* 1495–1502.

Piotrowski, C. (1996). Use of the Beck Depression Inventory in clinical practice. *Psychological Reports, 6,* 74–82.

Piotrowski, C. (1999). Assessment practices in the era of managed care: Current status and future directions. *Journal of Clinical Psychology, 55,* 787–796.

Piotrowski, C., & Keller, J. W. (1984). Psychodiagnostic testing in APA-approved clinical psychology programs. *Professional Psychology: Research and Practice, 15,* 450–456.

Piotrowski, C., & Keller, J. W. (1989). Psychological testing in outpatient mental health facilities: A national study. *Professional Psychology: Research and Practice, 20,* 423–425.

Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical training programs. *Journal of Personality Assessment, 61,* 394–405.

Piotrowski, Z. A. (1937). The Rorschach Ink-Blot Method in organic disturbances of the central nervous system. *Journal of Nervous and Mental Disorders, 86,* 525–537.

Piotrowski, Z. A. (1957). *Perceptanalysis.* New York: Macmillan.

Piotrowski, Z. A. (1960). The movement score. In M. Rickers-Ovsiankina (Ed.), *Rorschach psychology* (pp. 49–58). New York: Wiley.

Piotrowski, Z. A. (1969a). Long-term prognosis in schizophrenia based on Rorschach findings: The LTPTI. In D. V. Sira Sankar (Ed.), *Schizophrenia, current concepts and research* (pp. 84–103). Hicksville, NY: PJD.

Piotrowski, Z. A. (1969b). A Piotrowski interpretation. In J. E. Exner (Ed.), *The Rorschach systems.* New York: Grune & Stratton.

Piotrowski, Z. A., & Bricklin, B. A. (1961). A second validation of a long-term prognostic index for schizophrenic patients. *Journal of Consulting Psychology, 25,* 123–128.

Ponsford, J. L. (1988). Neuropsychological assessment: The need for a more pragmatic approach. *Australian Psychologist, 23,* 349–360.

Ponsford, J. L. (2000). Attention. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 355–400). New York: Wiley.

Ponsford, J. L., & Kinsella, G. (1992). Attentional deficits following closed head injury. *Journal of Clinical and Experimental Neuropsychology, 16,* 822–838.

Pope, K. S. (1992). Responsibilities in providing psychological test feedback to clients. *Psychological Assessment, 4,* 268–271.

Pope, K. S., Butcher, J. N., & Seelen, J. (2000). *The MMPI, MMPI-2 and MMPI-A in court: A practical guide for expert witnesses and attorneys* (2nd ed.). Washington, DC: American Psychological Association.

Porter, E. H. (1950). *An introduction to therapeutic counseling.* Boston: Houghton-Mifflin.

Preston, J., O'Neal, J. H., & Talaga, M. C. (2002). *Handbook of clinical psychopharmacology for therapists* (3rd ed.). Oakland, CA: New Harbinger.

Pretzer, J. L., Epstein, N., & Fleming, B. (1992). The Marital Attitude Survey: A measure of dysfunctional attitudes and expectancies. *Journal of Cognitive Psychotherapy, 5,* 131–148.

Prigatano, G. P. (1983). Neuropsychological test performance in mildly hypoxemic patients with chronic obstructive pulmonary disease. *Journal of Consulting and Clinical Psychology, 51,* 108–116.

Prigatano, G. P. (1986). *Neuropsychological rehabilitation after brain injury.* Baltimore: Johns Hopkins University Press.

Prigatano, G. P. (1987). Personality and psychosocial consequences after brain injury. In M. Meir, A. Benton, & L. Diller (Eds.), *Neuropsychological rehabilitation* (pp. 335–378). New York: Plenum Press.

Prigatano, G. P. (1992). Personality disturbance associated with traumatic brain injury. *Journal of Consulting and Clinical Psychology, 60,* 360–368.

Prigatano, G. P. (1999). Impaired awareness, finger tapping, and rehabilitation outcome after brain injury. *Rehabilitation Psychology, 44,* 145–159.

Prigatano, G. P., Amin, K., & Rosenstein, L. D. (1992a). *Manual for the BNI Screen for Higher Cerebral Functions.* Phoenix, AZ: Barrow Neurological Institute.

Prigatano, G. P., Amin, K., & Rosenstein, L. D. (1992b). Validity studies of the BNI Screen for Higher Cerebral Functions. *Barrow Neurological Institute Quarterly, 9,* 2–9.

Prince, J. P., & Heisser, L. J. (2000). *Essentials of career interest assessment.* New York: Wiley.

Prochaska, J. O. (2000). Change at differing stages. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes and practices for the 21st century* (pp. 109–127). New York: Wiley.

Prochaska, J. O., & DiClemente, C. C. (1984). *The transtheoretical approach: Crossing the traditional boundaries of therapy.* Homewood, IL: Dow Jones-Irwin.

Prochaska, J. O., & DiClemente, C. C. (1992). The transtheoretical approach. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (pp. 300–334). New York: Basic Books.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47,* 1102–1114.

Prochaska, J. O., Norcross, J. C., & DiClemente, C. C. (1994). *Changing for good.* New York: Morrow.

Prochaska, J. O., Rossi, J. S., & Wilcox, N. S. (1991). Change processes and psychotherapy outcome in integrative case research. *Journal of Psychotherapy Integration, 1,* 103–120.

Pruit, S. D., Klapow, J. C., Epping-Jordan, J. E., & Dresselhaus, T. R. (1999). Moving behavioral medicine to the front line: A model for the integration of behavioral and medical science in primary care. *Professional Psychology: Research and Practice, 29,* 230–236.

Pruitt, J. A., Smith, M. C., Thelen, M. H., & Lubin, B. (1985). Attitudes of academic clinical psychologists toward projective techniques: 1968–1983. *Professional Psychology: Research and Practice, 16,* 781–788.

Psychological Corporation. (1997). *WAIS-III/WMS-III technical manual.* San Antonio, TX: Psychological Corporation.

Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Author.

Psychological Corporation. (2002). *Technical manual for the WPPSI-III.* San Antonio, TX: Psychological Corporation.

Pugh, G. (1985). The California Psychological Inventory and police selection. *Journal of Police Science and Administration, 13,* 172–177.

Puig-Antich, J., & Chambers, W. (1978). *The schedule for affective disorders and schizophrenia for school aged children.* New York: New York State Psychiatric Institute.

Putzke, J. D., Williams, M. A., Daniel, F. J., & Boll, T. (1999). The utility of *K*-Correction to adjust for defensive response set on the MMPI. *Assessment, 6,* 61–70.

Query, W. T. (1966). CPI factors and success of seminary students. *Psychological Reports, 18,* 665–660.

Quillan, J., Besing, S., & Dinning, D. (1977). Standardization of the Rathus Assertiveness Schedule. *Journal of Clinical Psychology, 33,* 418–422.

Rao, S. M. (1990). Neuroimaging correlates of cognitive dysfunction. In S. M. Rao (Ed.), *Neurobehavioral aspects of multiple sclerosis* (pp. 118–135). New York: Oxford University Press.

Rapaport, C., Gill, M., & Schafer, J. (1968). *Diagnostic psychological testing* (Vol. 1). (Rev. ed.). Chicago: Year Book.

Rasmusson, D. X., Zonderman, A. B., Kawas, C., & Resnick, S. M. (1998). Effects of age and dementia on the Trail Making Test. *Clinical Neuropsychologist, 12,* 169–178.

Rathus, S. A. (1972). An experimental investigation of assertive training in a group setting. *Journal of Behavior Therapy and Experimental Psychiatry, 3,* 81–86.

Rathus, S. A. (1973). A 30-item schedule for assessing assertive behavior. *Behavior Therapy, 4,* 398–406.

Rathus, S. A., & Nevid, J. S. (1977). Concurrent validity of the 30-item assertiveness schedule with a psychiatric population. *Behavior Therapy, 8,* 393–397.

Rawls, J. R., & Slack, G. K. (1968). Artists versus non-artists: Rorschach determinants and artistic creativity. *Journal of Projective Techniques and Personality Assessment, 32,* 233–237.

Redfering, D. L., & Collings, J. (1982). A comparison of the Koppitz and Hutt techniques of Bender-Gestalt administration correlated with WISC-R performance scores. *Educational and Psychological Measurement, 42,* 41–47.

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment, 10,* 10–20.

Reich, J. H., & Noyes, R. (1987). A comparison of *DSM-III* personality disorders in acutely ill panic and depressed patients. *Journal of Anxiety Disorders, 1,* 123–131.

Reich, W. (2000). Diagnostic Interview for Children and Adolescents (DICA). *Journal of the Academy of Child and Adolescent Psychiatry, 39,* 59–66.

Reichenberg, N., & Raphael, A. R. (1992). *Advanced psychodiagnostic interpretation of the Bender Gestalt Test.* New York: Praeger.

Reid, D. B., & Kelly, M. P. (1993). Wechsler Memory Scale-Revised in closed head injury. *Journal of Clinical Psychology, 49,* 245–254.

Reinecke, M. A., Beebe, D. W., & Stein, M. A. (1999). The third factor of the WISC-III: Its (probably) not freedom from distractibility. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 322–328.

Reitan, R. M. (1955a). Certain differential effects of left and right cerebral lesions in human adults. *Journal of Comparative and Physiological Psychology, 48,* 474–477.

Reitan, R. M. (1955b). Validity of the Rorschach Test as a measure of the psychological effects of brain damage. *Archives of Neurology and Psychiatry, 73,* 445–451.

Reitan, R. M. (1974a). Methodological problems in clinical neuropsychology. In R. M. Reitan & L. A. Davison (Eds.), *Clinical neuropsychology: Current status and applications.* New York: Wiley.

Reitan, R. M. (1974b). Psychological effects of cerebral lesions in children of early school age. In R. M. Reitan & L. A. Davison (Eds.), *Clinical neuropsychology: Current status and applications* (pp. 53–90). Washington, DC: Winston & Sons.

Reitan, R. M., & Wolfson, D. (1985). *The Halsted-Reitan Neuropsychological Test Battery: Theory and clinical interpretation.* Tucson, AZ: Tucson Neuropsychology Press.

Reitan, R. M., & Wolfson, D. (1992). *Neuropsychological education of older children.* Tucson, AZ: Neuropsychology Press.

Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation* (2nd ed.). Tucson, AZ: Neuropsychology Press.

Retief, A. (1987). Thematic apperception testing across cultures: Tests of selection versus tests of inclusion. *South African Journal of Psychology, 17,* 47–55.

Retzlaff, P. D. (1995). *Tactical psychotherapy of the personality disorders: An MCMI-III-based approach.* Boston: Allyn & Bacon.

Retzlaff, P. D., Lorr, M., & Hyer, L. (1989). *An MCMI-II item-level component analysis: Personality and clinical factors.* Unpublished manuscript.

Retzlaff, P. D., Sheehan, E. P., & Fiel, A. (1991). MCMI-II report style and bias: Profile and validity scales analysis. *Journal of Personality Assessment, 56,* 478–486.

Retzlaff, P. D., & Dunn, T. (2003). The Millon Clinical Multiaxial Inventory. In L. E. Beutler & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality.* New York: Guilford Press.

Rey, A. (1941). Psychological examination of traumatic encephalopathy. *Archives de Psychologie, 28,* 286–340.

Rey, A. (1964). *The clinical exam in psychology.* Paris: Presses Universitaires de France.

Reynolds, C. R. (1986). Wide Range Achievement Test (WRAT-R), 1984 edition. *Journal of Counseling and Development, 64,* 540–541.

Reynolds, C. R., Chastain, R. L., Kaufman, A. S., & McLean, J. E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. *Journal of School Psychology, 25,* 323–324.

Reynolds, C. R., & Hartlage, L. (1979). Comparison of WISC and WISC-R regression lines for academic prediction with black and white referred children. *Journal of Consulting and Clinical Psychology, 47,* 589–591.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children.* Circle Pines, MN: American Guidance Service.

Ridgeway, E. M., & Exner, J. E. (1980). *Rorschach correlates of achievement needs in medical students under an arousal state.* Rorschach Workshops (Study No. 274, unpublished).

Riessman, F., & Miller, S. M. (1958). Social class and projective tests. *Journal of Projective Techniques, 22,* 432–439.

Riggs, D. S., & Foa, E. B. (1993). Obsessive compulsive disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (2nd ed., pp. 189–239). New York: Guilford Press.

Riskind, J. H., Beck, A. T., Berchick, R. J., Brown, G., & Steer, R. A. (1987). Reliability of *DSM-III* diagnoses for major depression and generalized anxiety disorder using the Structured Clinical Interview for *DSM-III. Archives of General Psychiatry, 44,* 817–820.

Ritzler, B. A., Sharkey, K. J., & Chudy, J. F. (1980). A comprehensive projective alternative to the TAT. *Journal of Personality Assessment, 44,* 358–362.

Robins, L. N., Cottler, L. B., Bucholz, K. K., & Compton, W. (1996). *The Diagnostic Interview Schedule (Version IV).* St. Louis, MO: Washington University School of Medical.

Robins, L. N., & Helzer, J. E. (1994). The half life of a structured interview: The NIMH Diagnostic Interview Schedule (DIS). *International Journal of Methods in Psychiatric Research, 4,* 95–102.

Robins, L. N., Helzer, J. E., Cottler, L. B., & Goldring, E. (1989). *NIMH Diagnostic Interview Schedule (Version III-Revised).* St. Louis, MO: Washington University School of Medicine.

Robins, L. N., Helzer, J. E., Croughan, J. L., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule. *Archives of General Psychiatry, 38,* 381–389.

Robinson, J. (2001). *Brain calipers: Descriptive psychopathology and the mental status examination* (2nd ed.). Port Huron, MI: Rapid Psychler Press.

Rodenhauser, P., & Fornal, R. E. (1991). How important is the mental status examination? *Psychiatric Hospital, 22,* 256–262.

Rodgers, D. A. (1972). The MMPI: A review. In O. K. Buros (Ed.), *Seventh mental measurements yearbook* (Vol. 1, pp. 243–250). Highland Park, NJ: Gryphon Press.

Roe, A. (1952). Analysis of group Rorschachs of psychologists and anthropologists. *Journal of Projective Techniques, 16,* 212–242.

Rogers, C. R. (1961). A process conception of psychotherapy. In C. R. Rogers (Ed.), *On becoming a person.* Boston: Houghton Mifflin.

Rogers, C. R. (1992). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting and Clinical Psychology, 60,* 827–832. (Original work published 1957)

Rogers, R. (1984). *Rogers Criminal Responsibility Assessment Scales.* Odessa, FL: Psychological Assessment Resources.

Rogers, R. (1995). *Diagnostic and structured interviewing: A handbook for psychologists.* Odessa, FL: Psychological Assessment Resources.

Rogers, R. (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.

Rogers, R., Bagby, R. M., & Chakraborty, D. (1993). Feigning schizophrenic disorders on the MMPI-2: Detection of coached simulators. *Journal of Personality Assessment, 60,* 215–226.

Rogers, R., Salekin, R. T., & Sewell, K. W. (1999). Validation of the Millon Clinical Multiaxial Inventory for Axis II disorders: Does it meet the *Daubert* standard? *Law and Human Behavior, 22,* 425–443.

Rogers, R., Sewell, K. W., & Ustad, K. L. (1995). Feigning among chronic outpatients on the MMPI-2: A systematic examination of fake-bad indicators. *Assessment, 2,* 81–89.

Roid, G. H., Prifitera, A., & Ledbetter, M. (1988). Confirmatory analysis of the factor of the Wechsler Memory Scale-Revised. *Clinical Neuropsychologist, 2,* 116–120.

Romanczyk, R. G., Kent, R. N., Diament, C., & O'Leary, K. D. (1973). Measuring the reliability of observational data: A reactive process. *Journal of Applied Behavior Analysis, 6,* 175–184.

Rorschach, H. (1941). *Psychodiagnostik* (Hans Huber Verlag, Trans.). Bern, Switzerland: Bircher. (Original work published 1921)

Rosen, B. M., Bahn, A. K., & Kramer, M. (1964). Demographic and diagnostic characteristics of psychiatric clinic patients in the U.S.A., 1961. *American Journal of Orthopsychiatry, 34,* 455–468.

Rosenhan, D. L. (1973). On being sane in insane places. *Science, 179,* 250–257.

Rosenthal, M. J. (1962). Some behavior correlated to the Rorschach experience-balance. *Journal of Projective Techniques, 26,* 442–446.

Rosenthal, M. J. (1989). Towards selective and improved performance of the mental status examination. *Acta Psychiatrica Scandinavica, 80,* 207–215.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R., & Fode, K. L. (1963). The effects of experimenter bias on the performance of the albino rat. *Behavioral Science, 8,* 183–189.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart and Winston.

Rosenzweig, S. (1976). *Manual for the Rosenzweig Picture-Frustration Study, adolescent form.* St. Louis, MO: Author.

Rosenzweig, S. (1977). *Manual for the Children's Form of the Rosenzweig Picture-Frustration (P-F) Study.* St. Louis, MO: Rana House.

Rosenzweig, S. (1978). *Adult form supplement to the Basic Manual of the Rosenzweig Picture Frustration (P-F) Study.* St. Louis, MO: Rana House.

Ross, L. D. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press.

Rossi, A., & Neuman, G. (1961). A comparative study of Rorschach norms: Medical students. *Journal of Projective Techniques, 25,* 334–338.

Rossini, E. D., & Kaspar, J. C. (1987). The validity of the Bender-Gestalt emotional indicators. *Journal of Personality Assessment, 51,* 254–261.

Rossini, E. D., & Moretti, R. J. (1997). Thematic Apperception Test (TAT) interpretation: Practice recommendations from a survey of clinical psychology doctoral programs by the American Psychological Association. *Professional Psychology: Research and Practice, 28,* 393–398.

Roth, D. L., Hughes, C. W., Mankowski, P. G., & Crosson, B. (1984). Investigation of validity of WAIS-R short forms for patients suspected to have brain impairment. *Journal of Consulting and Clinical Psychology, 52,* 722–723.

Rouse, S. V., Butcher, J. N., & Miller, K. B. (1999). Assessment of substance abuse in psychotherapy clients: The effectiveness of the MMPI-2 substance abuse scales. *Psychological Assessment, 11,* 101–107.

Royer, F. L., & Holland, T. R. (1975). Rotations of visual design in psychopathological groups. *Journal of Consulting and Clinical Psychology, 43,* 346–356.

Rude, S. R. (1986). Relative benefits of assertion or cognitive self-control treatment for depression as a function of proficiency in each domain. *Journal of Consulting and Clinical Psychology, 54,* 390–394.

Ruegg, R. G., Ekstrom, D. E., Evans, D. L., & Golden, R. N. (1990). Introduction of a standardized report form improves the quality of mental status examination reports by psychiatric residents. *Academic Psychiatry, 14,* 157–163.

Ruff, R. M., & Parker, S. B. (1993). Gender and age-specific changes in motor speed and eye-hand coordination in adults: Normative values for the Finger Tapping and Grooved Pegboard Tests. *Perceptual and Motor Skills, 76,* 1219–1230.

Ruscio, J. (2000). The role of complex thought in clinical prediction: Social accountability and the need for cognition. *Journal of Consulting and Clinical Psychology, 68,* 145–154.

Rushton, J. P. (1994). The equalitarian dogma revisited. *Intelligence, 19,* 263–280.

Russ, S. W. (2001). Tackling ethical dilemmas in personality assessment. *Journal of Personality Assessment, 77,* 255–258.

Russell, E. W. (1975). A multiple scoring method for the assessment of complex memory functions. *Journal of Consulting and Clinical Psychology, 43,* 800–809.

Russell, E. W. (1979). Three patterns of brain damage on the WAIS. *Journal of Clinical Psychology, 35,* 611–620.

Russell, E. W. (1988). Renorming Russell's version of the Wechsler Memory Scale. *Journal of Clinical and Experimental Neuropsychology, 10,* 235–249.

Russell, E. W. (2000). The cognitive-metric, fixed battery approach to neuropsychological assessment. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 449–483). Hillsdale, NJ: Erlbaum.

Ryan, C. M., Morrow, L., Parklinson, D., & Branet, E. (1987). Low level lead exposure and neuropsychological functioning in blue collar males. *International Journal of Neuroscience, 36,* 29–39.

Ryan, J. J. (1999). Two types of tables for use with the seven-subtest short forms of the WAIS-III. *Journal of Psychoeducational Assessment, 17,* 145–151.

Ryan, J. J., Georgemiller, R., & McKinney, B. (1984). Application of the four-subtest WAIS-R short form with an older clinical sample. *Journal of Clinical Psychology, 40,* 1033–1036.

Ryan, J. J., Lopez, S. J., & Werth, T. R. (1998). Administration time estimates for WAIS-III subtests, scales, and short forms in a clinical sample. *Journal of Psychoeductional Assessment, 16,* 315–323.

Ryan, J. J., Lopez, S. J., & Werth, T. R. (1999). Development and preliminary validation of a Satz-Mogel short form of the WAIS-III in a sample of persons with substance abuse disorders. *International Journal of Neurosciences, 98,* 131–140.

Ryan, J. J., Paolo, A. M., & Brungardt, T. M. (1990). Standardization of the Wechsler Adult Intelligence Scale-Revised for persons 75 years and older. *Psychological Assessment, 2,* 404–411.

Ryan, J. J., & Paul, C. A. (1999). Who is president of the United States? *Perceptual and Motor Skills, 89,* 595–596.

Ryan, J. J., Paul, C. A., & Arb, J. D. (1999). Intrasubtest scatter on the WAIS-III Information subtest and psychometrically defined retrieval deficits. *Perceptual and Motor Skills, 89,* 1052–1058.

Ryan, J. J., Rosenberg, S. J., & Mittenberg, W. (1984). Factor analysis of the Rey Auditory Verbal Learning Test. *International Journal of Clinical and Experimental Neuropsychology, 5,* 249–253.

Ryan, J. J., & Ward, L. C. (1999). Validity, reliability, and standard errors of measurement for two seven subtest short forms of the Wechsler Adult Intelligence Scale-III. *Psychological Assessment, 11,* 207–211.

Rzechorzek, A. (1979). Cognitive dysfunctions resulting from unilateral frontal lesions in man. In M. Molloy, G. V. Stanley, & K. W. Walsh (Eds.), *Brain impairment: Proceedings of the 1978 Brain Impairment Workshop.* Melbourne: University of Melbourne, Australia.

Saklofske, D. H., Hildebrand, D. K., & Gorsuch, R. L. (2000). Replication of the factor structure of the Wechsler Adult Intelligence Scale-Third Edition with a Canadian sample. *Psychological Assessment, 12*(4), 436–439.

Sales, J., & Miller, P. (1994). *Psychology in litigation and legislation.* Washington, DC: American Psychological Association.

Salmon, R., Arnold, J. M., & Collyer, Y. M. (1972). What do the determinants determine? The internal validity of the Rorschach. *Journal of Personality Assessment, 36,* 33–38.

Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development, 64,* 80–97.

Sammons, M. T., & Schmidt, N. B. (2001). *Combined treatment for mental disorders: A guide to psychological and pharmacological interventions.* Washington, DC: American Psychological Association.

Sandoval, J., Frisby, C. L., Geisinger, K. F., Ramos-Grenier, J., & Scheuneman, J. D. (1999). *Test interpretation and diversity: Achieving equity in assessment.* Washington, DC: American Psychological Association.

Sandy, L. R. (1986). The descriptive-collaborative approach to psychological report writing. *Psychology in the Schools, 23,* 395–400.

Sarason, I. G., Levine, H. M., Basham, R. B., & Sarason, B. R. (1983). Assessing social support: The Social Support Questionnaire. *Journal of Personality and Social Psychology, 44,* 127–139.

Sarchione, C. D., Cuttler, M. J., Muchinsky, P. M., & Nelson-Gray, R. O. (1998). Prediction of dysfunctional job behaviors among law enforcement officers. *Journal of Applied Psychology, 83,* 904–912.

Sarwer, D. B., & Sayers, S. L. (1998a). Assessment of marital dysfunction. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 293–314). Boston: Allyn & Bacon.

Sarwer, D. B., & Sayers, S. L. (1998b). Behavioral interviewing. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 63–78). Boston: Allyn & Bacon.

Sattler, J. M. (1973a). Examiners scoring style, accuracy, ability, and culturally disadvantaged children. In L. Mann & D. Sabatino (Eds.), *The first review of special education* (Vol. 2). Philadelphia: J. S. E. Press.

Sattler, J. M. (1973b). Racial experimenter effects. In K. S. Miller & R. M. Dreger (Eds.), *Comparative studies of blacks and whites in the United States* (pp. 8–32). New York: Seminar Press.

Sattler, J. M. (1985). Review of the Hutt adaptation of the Bender-Gestalt Test. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (Vol. 1, pp. 184–185). Highland Park, NJ: Gryphon Press.

Sattler, J. M. (1992). *Assessment of children* (3rd ed. rev.). San Diego, CA: Author.

Sattler, J. M. (2001). *Assessment of children: Cognitive functions* (4th ed.). San Diego, CA: Author.

Sattler, J. M. (2002). *Assessment of children: Behavioral and clinical applications* (4th ed.). San Diego, CA: Author.

Sattler, J. M., & Atkinson, L. (1993). Item equivalence across scales: The WPPSI-R and WISC-III. *Psychological Assessment, 5,* 203–206.

Sattler, J. M., & Gwynne, J. (1982). White examiners generally do not impede the intelligence test performance of black children: To debunk a myth. *Journal of Consulting and Clinical Psychology, 50,* 196–208.

Sattler, J. M., Hillix, W. A., & Neher, L. A. (1970). Halo effect in examiner scoring of intelligence test responses. *Journal of Consulting and Clinical Psychology, 34,* 172–176.

Sattler, J. M., & Ryan, J. J. (1998). *Assessment of children, revised and updated 3rd ed. WAIS-III supplement.* San Diego, CA: Jerome Sattler.

Sattler, J. M., & Winget, B. M. (1970). Intelligence testing procedures as affected by expectancy and I.Q. *Journal of Clinical Psychology, 26,* 446–448.

Satz, P., & Mogel, S. (1962). An abbreviation of the WAIS for clinical use. *Journal of Clinical Psychology, 18,* 77–79.

Savage, R. M., & Gouvier, W. D. (1992). Rey Auditory-Verbal Learning Test: The effects of age and gender, and norms for delayed recall and story recognition trials. *Archives of Clinical Neuropsychology, 1,* 407–414.

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing. *American Psychologist, 40,* 355–366.

Sbordone, R. J. (2000a). The assessment interview in clinical neuropsychology. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 94–128). New York: Wiley.

Sbordone, R. J. (2000b). The executive functions of the brain. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 437–456). New York: Wiley.

Sbordone, R. J., & Guilmette, T. J. (1999). Ecological validity: Prediction of everyday and vocational functioning from neuropsychological test data. In J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice* (pp. 223–250). New York: Swets.

Sbordone, R. J., & Long, C. J. (Eds.). (1996). *Ecological validity of neuropsychological testing.* Odessa, FL: Psychological Assessment Resources.

Schachtel, E. G. (1966). *Experimental foundations of Rorschach's test.* New York: Basic Books.

Schafer, R. (1954). *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton.

Schalock, R. L., Stark, J. A., Snell, M. E., Coulter, D. L., Polloway, E. A., Luckasson, R., et al. (1994). The changing conception of mental retardation: Implications for the field. *Mental Retardation, 32,* 181–193.

Schinka, J. A. (1983). *Neuropsychological Status Examination.* Odessa, FL: Psychological Assessment Resources.

Schinka, J. A., LaLone, L., & Greene, R. (1998). Effects of psychopathology and demographic characteristics on MMPI-2 scale scores. *Journal of Personality Assessment, 70,* 197–211.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627–670.

Schmidt, H. O., & Fonda, C. P. (1954). Rorschach scores in the manic state. *Journal of Psychology, 38,* 427–437.

Schmidt, M. (1996). *Rey Auditory Verbal Learning Test.* Los Angeles: Western Psychological Services.

Schmidt, M., Trueblood, W., Merwin, M., & Durham, R. L. (1994). How much do "attention" tests tell us? *Archives of Clinical Neuropsychology, 9,* 383–394.

Schmitz, N., & Hartkamp, N. (2000). Assessing clinically significant change: Application to the SCL-90-R. *Psychological Reports, 86,* 263–274.

Schmitz, N., Kruse, J., Heckrath, L., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: The General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology, 34,* 360–366.

Schraa, J. C., Jones, N. F., & Dirks, J. E. (1983). Bender-Gestalt recall: A review of the normative data and related issues. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 2, pp. 125–138). Hillsdale, NJ: Erlbaum.

Schroth, M. L. (1987). Relationships between achievement-related motives, extrinsic conditions, and task performance. *Journal of Social Psychology, 127,* 39–48.

Schuerger, J. M., & Witt, A. C. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology, 45,* 294–302.

Schultheiss, O. C., & Brunstein, J. C. (2001). Assessment of implicit motives with a research version of the TAT: Picture profiles, gender differences, and relations to other personality measures. *Journal of Personality Assessment, 77,* 71–86.

Schultz, C. B., & Sherman, R. H. (1976). Social class, development, and differences in reinforcer effectiveness. *Review of Educational Research, 46,* 25–59.

Schwab-Stone, M., Fisher, P., Piacentini, J., Shaffer, D., Davies, M., & Briggs, M. (1993). The Diagnostic Interview Schedule for Children-Revised version (DISC-R). II: Test-retest reliability. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 651–657.

Schwartz, G. E. (1982). Testing the biopsychosocial model: The ultimate challenge facing behavioral medicine? *Journal of Consulting and Clinical Psychology, 50,* 1040–1053.

Schwartz, L., & Levitt, E. E. (1960). Short forms of the WISC for children in the educable, non-institutionalized mentally retarded. *Journal of Educational Psychology, 51,* 187–190.

Schwartz, R. M., & Garamoni, G. L. (1989). Cognitive balance and psychopathology: Evaluation of an information processing model of positive and negative states of mind. *Clinical Psychology Review, 9,* 271–294.

Schwartz, S., & Wiedel, T. C. (1981). Incremental validity of the MMPI in neurological decision-making. *Journal of Personality Assessment, 45,* 424–426.

Seeman, K., Yesavage, J., & Widrow, L. (1985). Correlations of self-directed violence in acute schizophrenics with clinical ratings and personality measures. *Journal of Nervous and Mental Diseases, 173,* 298–302.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50,* 965–974.

Seligman, M. E. P., Abramson, L. Y., Semmel, A., & von Baeyer, C. (1979). Depressive attributional style. *Journal of Abnormal Psychology, 88,* 242–247.

Sewitch, T., & Kirsch, I. (1984). The cognitive content of anxiety: Naturalistic evidence for the predominance of threat-related thoughts. *Cognitive Therapy and Research, 8,* 49–58.

Shaffer, D., Fisher, P. W., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children (Version IV; NIMH DISC-IV):

Description, differences from previous versions, and reliability of some common diagnosis. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 28–38.

Shaffer, D., Schwab-Stone, M., Fisher, P., Cohen, P., Piacentini, J., Davies, M., et al. (1993). The Diagnostic Interview Schedule for Children-Revised version (DISC-R). I: Preparation, field testing, interrater reliability, and acceptability. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 643–650.

Shalit, B. (1965). Effects of environmental stimulation on the M, FM, and m response to the Rorschach. *Journal of Projective Techniques and Personality Assessment, 29,* 228–231.

Shapiro, D. A. (1960). Perceptual understanding of color response. In M. Rickers-Ovsiankina (Ed.), *Rorschach psychology* (pp. 89–103). New York: Wiley.

Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 62,* 522–534.

Shapiro, D. M., & Harrison, D. W. (1990). Alternative forms of the AVLT: A procedure and test of form equivalency. *Archives of Clinical Neuropsychology, 5,* 405–410.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156.

Shapiro, S. K., & Simpson, R. G. (1995). Koppitz scoring system as a measure of Bender-Gestalt performance in behaviorally disturbed adolescents. *Journal of Clinical Psychology, 51,* 108–112.

Sharkey, K. J., & Ritzler, B. A. (1985). Comparing diagnostic validity of the TAT and a new Picture Projective Test. *Journal of Personality Assessment, 49,* 406–412.

Shatin, L. (1952). Psychoneurosis and psychosomatic reactions: A Rorschach contrast. *Journal of Consulting Psychology, 16,* 220–223.

Shaw, D. S., & Gynther, M. D. (1986). An attempt to obtain configural correlates for the California Psychological Inventory. *Psychological Reports, 59,* 675–678.

Shear, M. K., Greeno, C., Kang, J., Ludewig, D., Frank, E., Swartz, H. A., et al. (2000). Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry, 157,* 581–587.

Sherman, E. M. S., Strauss, E., Slick, D. J., & Spellacy, F. (2000). Effect of depression on neuropsychological functioning in head injury: Measurable but minimal. *Brain Injury, 14,* 621–632.

Shneidman, E. S. (Ed.). (1951). *Thematic test analysis.* New York: Grune & Stratton.

Shorkey, C. L., Reyes, E., & Whiteman, V. L. (1977). Development of the rational behavior inventory: Initial validity and reliability. *Educational and Psychological Measurement, 37,* 527–534.

Shrauger, J. S., & Osberg, T. M. (1981). The relative accuracy of self-predictions and judgments of others in psychological assessment. *Psychological Bulletin, 90,* 322–351.

Sieber, K. O., & Meyers, L. S. (1992). Validation of the MMPI-2 social introversion subscales. *Psychological Assessment, 4,* 185–189.

Siegel, A. W., Schechter, M. D., & Diamond, S. P. (1996). Neurobehavioral assessment format. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 429–504). Delray Beach, FL: St. Lucie Press.

Siegel, L. S. (1999). Issues in the definition and diagnosis of learning disabilities: A perspective on *Guckenberger v. Boston University. Journal of Learning Disabilities, 32,* 304–319.

Silverstein, A. B. (1990). Short forms of individual intelligence tests. *Psychological Assessment, 2,* 3–11.

Silverstein, A. B., & Mohan, P. J. (1962). Bender-Gestalt figure rotations in the mentally retarded. *Journal of Consulting Psychology, 26,* 386–388.

Simon, W. E. (1969). Expectancy effects in the scoring of vocabulary items: A study of scorer bias. *Journal of Educational Measurement, 6,* 159–164.

Simons, A. D., & Thase, M. E. (1992). Biological markers, treatment outcome, and 1 year follow-up in endogenous depression: Electroencephalographic sleep studies and response to cognitive therapy. *Journal of Consulting and Clinical Psychology, 60,* 392–401.

Singer, J. L. (1960). The experience type: Some behavioral correlates and theoretical implications. In M. Rickers-Ovsiankina (Ed.), *Rorschach psychology.* New York: Wiley.

Singer, R. M. (1990). *Neurotoxicity guidebook.* New York: Van Nostrand-Reinhold.

Sipps, G. J., Berry, G. W., & Lynch, E. M. (1987). WAIS-R and social intelligence: A test of established assumptions that uses the CPI. *Journal of Clinical Psychology, 43,* 499–504.

Skolnick, A. (1966). Motivational imagery and behavior over twenty years. *Journal of Consulting Psychology, 30,* 463–478.

Slate, R. J., & Hunnicutt, L. C. (1988). Examiner errors on the Wechsler scales. *Journal of Psychoeducational Assessment, 6,* 280–288.

Slate, R. J., Jones, C. H., & Murray, R. A. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale-Revised: An empirical evaluation of practice administrations. *Professional Psychology, 22,* 375–379.

Smith, A. (1982). *Symbol Digit Modalities Test (SDMT). Manual* (Rev.). Los Angeles: Western Psychological Services.

Smith, C. P., & Graham, J. R. (1981). Behavioral correlates for the MMPI standard F scale and the modified F scale for black and white psychiatric patients. *Journal of Consulting and Clinical Psychology, 49*(3), 455–459.

Smith, D., & Dumont, F. (1995). A cautionary study: Unwarranted interpretations of the Draw-A-Person Test. *Professional Psychology: Research and Practice, 26,* 298–303.

Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7,* 300–308.

Smith, M. A., & Senior, C. (2001). The Internet and clinical psychology: A general review of the implications. *Professional Psychology, 21,* 129–136.

Smith, M. L., Glass, G. V., & Miller, T. L. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press.

Smith, N. M. (1981). The relationship between the Rorschach whole response and level of cognitive functioning. *Journal of Personality Assessment, 45,* 13–19.

Smith, S. R., Baity, M. R., Knowles, E. S., & Hilsenroth, M. J. (2002). Assessment of disordered thinking in children and adolescents: The Rorschach Perceptual-Thinking Index. *Journal of Personality Assessment, 76,* 333–351.

Smith, S. R., Hilsenroth, M. J., Castlebury, F. D., & Durham, T. W. (1999). The clinical utility of the MMPI-2 Antisocial Practices Content Scale. *Journal of Personality Disorders, 13,* 385–393.

Snow, W. G., Tierney, M. C., Zorzitto, M. L., Fisher, R. H., & Reid, D. W. (1988). One year test-retest reliability of selected tests in older adults [Abstract]. *Journal of Clinical and Experimental Neuropsychology, 10,* 60.

Snyder, P. J., & Nussbaum, P. D. (Eds.). (1998). *Clinical neuropsychology: A pocket handbook for assessment.* Washington, DC: American Psychological Association.

Snyder, W. V. (1945). An investigation of the nature of nondirective psychotherapy. *Journal of General Psychology, 33,* 139–223.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42,* 137–144.

Sobel, D. S. (2000). Mind matters, money matters: The cost effectiveness of mind-body medicine. *Journal of the American Medical Association, 284,* 1705.

Sokolowski, K., Schmalt, H.-D., Langens, T., & Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: The Multi-Motive Grid (MMG). *Journal of Personality Assessment, 74,* 126–145.

Sommer, R., & Sommer, D. (1958). Assaultiveness and two types of Rorschach color responses. *Journal of Consulting Psychology, 22,* 57–62.

Sommers-Flanagan, R., & Sommers-Flanagan, J. (1999). *Clinical interviewing* (2nd ed.). New York: Wiley.

Southern, M. L., & Plant, W. T. (1968). Personality characteristics of very bright adults. *Journal of Social Psychology, 75,* 119–126.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112,* 140–154.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Spielberger, C. D. (1973). *Manual for the State-Trait Anxiety Inventory for Children.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., & Piotrowski, C. (1990). Clinician's attitudes toward computer-based testing. *Clinical Psychologist, 43,* 60–63.

Spielberger, C. D., Sydeman, S. J., Owen, A. E., & Marsh, B. J. (1999). Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (2nd ed.). Mahwah, NJ: Erlbaum.

Spiker, D. G., & Ehler, J. G. (1984). Structured psychiatric interviews for adult. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (pp. 291–304). New York: Pergamon Press.

Spirito, A., Faust, D., Myers, B., & Bechtel, D. (1988). Clinical utility of the MMPI in the evaluation of adolescent suicide attempters. *Journal of Personality Assessment, 52,* 204–211.

Spitzer, R. L., Endicott, J., & Cohen, J. (1974). Constraints on the validity of computer diagnosis. *Archives of General Psychiatry, 31,* 197–203.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry, 35,* 773–782.

Spitzer, R. L., Williams, J. B. W., & Gibbon, M. (1987). *Structured clinical interview for DSM-III-R (SCID).* New York: State Psychiatric Institute.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *Structured clinical interview for DSM-III-R personality disorders (SCID-II).* Washington, DC: American Psychiatric Press.

Spreen, O., & Benton, A. L. (1977). *Neurosensory Center Comprehensive Examination for Aphasia (NCCEA)* (Rev. ed.). Victoria, B.C., Canada: University of Victoria, Neuropsychology Laboratory.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests.* New York: Oxford University Press.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Oxford University Press.

Springer, S. P., & Deutsch, G. (1998). *Left brain right brain: Perspectives from cognitive neuroscience* (5th ed.). New York: Freeman.

Standage, K. (1986). Socialization scores in psychiatric patients and their implications for the diagnosis of personality disorders. *Canadian Journal of Psychiatry, 31,* 138–141.

Standage, K. (1990). A classification of respondents to the CPI Socialization scale: Associations with diagnosis and other clinical variables. *Personality and Individual Differences, 11,* 335–341.

Standage, K., Smith, D., & Norman, R. (1988). A classification of respondents to the CPI Socialization Scale: Associations with psychiatric diagnosis and implications for research. *Personality and Individual Differences, 9,* 231–236.

Stanley, M. A., Beck, J. G., & Zebb, B. J. (1996). Psychometric properties of four anxiety measures in older adults. *Behaviour Research and Therapy, 34,* 827–838.

Stanton, H. C., & Reynolds, C. R. (1998). *Configural frequency analysis as a method of determining Wechsler profile types.* Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Starcevic, V., Bogojevic, G., & Marinkovic, J. (2000). The SCL-90-R as a screening instrument for severe personality disturbance among outpatients with mood and anxiety disorders. *Journal of Personality Disorders, 14,* 199–207.

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Clark, R. P., Mulvey, E. P., et al. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior, 24,* 83–100.

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology, 55,* 117–118.

Steer, R. A., Beck, A. T., & Garrison, B. (1986). Applications of the Beck Depression Inventory. In N. Sartorius & T. A. Ban (Eds.), *Assessment of depression* (pp. 121–142). Geneva, Switzerland: World Health Organization.

Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory-II with adolescent depressed outpatients. *Journal of Psychopathology and Behavioral Assessment, 20,* 127–137.

Steer, R. A., Rissmiller, D. J., & Beck, A. T. (2000). Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behaviour Research and Therapy, 38,* 311–318.

Steichen, E. (1955). *Family of man.* New York: Simon & Schuster.

Stein, L. A. R., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Using the MMPI-2 to detect substance abuse in an outpatient mental health setting. *Psychological Assessment, 11,* 94–100.

Stein, L. A. R., Graham, J. R., & Williams, C. L. (1995). Detecting fake-bad MMPI-A profiles. *Journal of Personality Assessment, 65,* 415–427.

Stein, M. (1981). *The Thematic Apperception Test: An introductory manual for its clinical use with adults* (2nd ed.). Springfield, IL: Charles C Thomas.

Steinberg, M. (1993). *Interviewer's guide to the Structured Clinical Interview for DSM-IV Dissociative Disorders (SCID-D).* Washington, DC: American Psychiatric Press.

Steiner, J. L., Tebes, J. K., Sledge, W. H., & Walker, M. L. (1995). A comparison of the structured clinical interview for *DSM-III-R* and clinical diagnosis. *Journal of Nervous and Mental Diseases, 183,* 365–369.

Steisel, I. M. (1952). The Rorschach Test and suggestibility. *Journal of Abnormal and Social Psychology, 47,* 607–614.

Steketee, G. (1994). Behavioral assessment and treatment planning with obsessive compulsive disorder: A review emphasizing clinical application. *Behavior Therapy, 25,* 613–633.

Stelmachers, Z. T. (1995). Assessing suicidal clients. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 367–379). New York: Oxford University Press.

Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology, 3,* 292–316.

Sternberg, R. J., & Kaufman, J. C. (1998). Innovation and intelligence testing: The curious case of the dog that didn't bark. *European Journal of Psychological Assessment, 12,* 175–182.

Sternberg, R. J., & Lubant, T. I. (1996). Investing in creativity. *American Psychologist, 51,* 677–688.

Stewart, L. H. (1962). Social and emotional adjustment during adolescence as related to the development of psychosomatic illness in adulthood. *Genetic Psychology Monographs, 65,* 175–215.

Stoloff, M. L., & Couch, J. V. (Eds.). (1992). *Computer use in psychology: A directory of software* (3rd ed.). Washington, DC: American Psychological Association.

Stompe, T., Ortwein-Swoboda, G., Strobl, R., & Freidman, A. (2000). The age of onset of schizophrenia and the theory of anticipation. *Psychiatry Research, 93,* 125–134.

Storandt, M., Botwinick, J., & Danzinger, W. L. (1986). Longitudinal changes: Patients with mild SDAT and matched health controls. In L. W. Poon (Ed.), *Handbook for clinical memory assessment of older adults.* Washington, DC: American Psychological Association.

Stormant, C. T., & Finney, B. C. (1953). Projection and behavior: A Rorschach study of assaultive mental hospital patients. *Journal of Projective Techniques, 17,* 349–360.

Stotsky, B. A. (1952). A comparison of remitting and non-remitting schizophrenics on psychological tests. *Journal of Abnormal and Social Psychology, 47,* 489–496.

Stout, C. E. (1997). *Psychological assessment in managed care.* New York: Wiley.

Stout, C. E., & Cook, L. P. (1999). New areas for psychological assessment in general health care settings: What to do today to prepare for tomorrow. *Journal of Clinical Psychology, 55,* 797–812.

Strack, S. (1999). Millon's normal personality styles and dimensions. *Journal of Personality Assessment, 72,* 426–436.

Streiner, D. L., & Miller, H. R. (1990). Maximum likelihood estimates of the accuracy of four diagnostic techniques. *Educational and Psychological Measurement, 50,* 653–662.

Stringer, A. Y., & Nadolne, M. J. (2000). Neuropsychological assessment: Contexts for contemporary clinical practice. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 26–47). New York: Wiley.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Strupp, H. H. (1958). The psychotherapist's contribution to the treatment process. *Behavioral Science, 3,* 34–67.

Stukenberg, K. W., Brady, C., & Klinetob, N. (2000). Psychiatric inpatients and the MMPI-2: Providing benchmarks. *Journal of Clinical Psychology,* 747–756.

Stukenberg, K. W., Dura, J. R., & Kiecolt-Glaser, J. K. (1990). Depression screening scale validation in an elderly, community-dwelling population. *Psychological Assessment, 2,* 134–138.

Sturgis, E. T., & Gramling, S. E. (1998). Psychophysiological assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 126–178). Boston: Allyn & Bacon.

Stuss, D. T., Bisschop, S. M., Alexander, M. P., Levine, B., Katz, D., & Izukawa, D. (2001). The Trail Making Test: A study in focal lesion patients. *Psychological Assessment, 13,* 230–239.

Stuss, D. T., Gow, C. A., & Hetherington, C. R. (1992). No longer Gage: Frontal lobe dysfunction and emotional changes. *Journal of Consulting and Clinical Psychology, 60,* 349–359.

Stuss, D. T., Stethem, L. L., Hugenholtz, H., & Richard, M. T. (1989). Traumatic brain injury: A comparison of three clinical tests, and analysis of recovery. *Clinical Neuropsychologist, 3,* 145–156.

Sullivan, K. (2000). Examiners errors on the Wechsler Memory Scale-Revised. *Psychological Reports, 87,* 234–240.

Sullivan, K., & Bowden, S. C. (1997). Which tests do neuropsychologists use? *Journal of Clinical Psychology, 53,* 657–661.

Summerfeldt, L. J., & Antony, M. M. (2002). Structured and semistructured diagnostic interviews. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 3–37). New York: Guilford Press.

Sundberg, N. D. (1955). The acceptability of "fake" versus "bona fide" personality test interpretations. *Journal of Abnormal and Social Psychology, 50,* 145–147.

Suzuki, L. A., Meller, P. J., & Ponterotto, J. G. (Eds.). (1996). *Handbook of multicultural assessment: Clinical, psychological, and educational applications.* Englewood Cliffs, NJ: Prentice-Hall.

Swartz, M. S., Blazer, D. G., George, L. K., Winfield, I., Zakris, J., & Dye, E. (1989). Identification of borderline personality with the NIMH Diagnostic Interview Schedule. *American Journal of Psychiatry, 146,* 200–205.

Sweet, J. J., Moberg, P. J., & Suchy, Y. (2000). Ten-year follow-up survey of clinical neuropsychologists. Part I: Practices and beliefs. *Clinical Neuropsychologist, 14,* 18–37.

Swensen, W. M., Pearson, J. S., & Osborne, D. (1973). *An MMPI source book: Basic item, scale, and pattern data on 50,000 medical patients.* Minneapolis: University of Minnesota Press.

Szasz, T. (1987). Justifying coercion through religion and psychiatry. *Journal of Humanistic Psychology, 27,* 158–174.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52,* 1–23.

Tallent, N. (1992). *The practice of psychological assessment.* Englewood Cliffs, NJ: Prentice-Hall.

Tallent, N. (1993). *Psychological report writing* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Talley, J. L. (1990). *Children's Auditory Verbal Learning Test.* Odessa, FL: Psychological Assessment Resources.

Talley, P. F., Strupp, H. S., & Morey, L. C. (1990). Matchmaking in psychotherapy: Patient-therapist dimensions and their impact on outcome. *Journal of Consulting and Clinical Psychology, 58,* 182–188.

Taylor, R. L., Kaufman, D., & Partenio, I. (1984). The Koppitz developmental scoring system for the Bender-Gestalt: Is it developmental? *Psychology in the Schools, 21,* 425–428.

Taylor, R. L., Sternberg, L., & Partenio, I. (1986). Performance of urban and rural children on the SOMPA: Preliminary investigation. *Perceptual and Motor Skills, 63,* 1219–1223.

Teevan, R. C., Diffenderfer, D., & Greenfield, N. (1986). Need for achievement and sociometric status. *Psychological Reports, 58,* 446.

Tellegen, A., & Ben-Porath, Y. S. (1992). The new uniform T scores for the MMPI-2: Rationale, derivation, and appraisal. *Psychological Assessment, 4,* 145–155.

Tellegen, A., & Ben-Porath, Y. S. (1993). Code-type comparability of the MMPI and MMPI-2: Analysis of recent findings and criticisms. *Journal of Personality Assessment, 61,* 489–500.

Temp, G. (1971). Test bias: Validity of the SAT for blacks and whites in thirteen integrated institutions. *Journal of Educational Measurement, 8,* 245–251.

Teng, E. L., Chui, H. C., & Saperia, D. (1990). Senile dementia: Performance on a neuropsychological test battery. *Recent advances in Cardiovascular Disease, 11,* 27–34.

Terman, L. M. (1916). *The measurement of intelligence.* Boston: Houghton Mifflin.

Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity.* New York: McGraw-Hill.

Terrell, F., Taylor, J., & Terrell, S. L. (1978). Effects of types of social reinforcement on the intelligence test performance of lower-class black children. *Journal of Consulting and Clinical Psychology, 46,* 1538–1539.

Thelen, M. H., Farmer, J., Wonderlich, S., & Smith, M. (1991). A revision of the Bulimia Test: The BULIT-R. *Psychological Assessment, 3,* 119–124.

Thiesen, J. W. (1952). A pattern analysis of structural characteristics of the Rorschach Test in schizophrenia. *Journal of Consulting Psychology, 16,* 365–370.

Thomas, A. D., & Dudek, S. Z. (1985). Interpersonal affect in Thematic Apperception Test responses: A scoring system. *Journal of Personality Assessment, 49,* 30–36.

Thomas, C. B., & Duszynski, K. R. (1985). Are words of the Rorschach predictors of disease and death? The case of "whirling." *Psychosomatic Medicine, 47,* 201–211.

Thompson, G. M. (1948). MMPI correlates of movement responses on the Rorschach. *American Psychologist, 3,* 348–349.

Thompson, L. L., Heaton, R. K., Mathews, C. G., & Grant, I. (1987). Comparison of preferred and nonpreferred hand performance on four neuropsychological tasks. *Clinical Neuropsychologist, 1,* 324–334.

Thorndike, R. L. (1959). The California Psychological Inventory: A review. In O. K. Buros (Ed.), *Fifth mental measurements yearbook.* Highland Park, NJ: Gryphon Press.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*(1).

Tierney, M. C., Nores, A., Snow, W. G., Fisher, R. H., Zorzitto, M. L., & Reid, D. W. (1994). Use of the Rey Auditory Verbal Learning Test in differentiating normal aging from Alzheimer's and Parkinson's dementia. *Psychological Assessment, 6,* 129–134.

Timbrook, R. E., & Graham, J. R. (1994). Ethnic differences on the MMPI-2. *Psychological Assessment, 6,* 212–217.

Todd, A. L., & Gynther, M. D. (1988). Have MMPI Mf scale correlates changed in the past 30 years? *Journal of Clinical Psychology, 44,* 505–510.

Todd, J., Coolidge, F., & Satz, P. (1977). The Wechsler Adult Intelligence Scale Discrepancy Index: A neuropsychological evaluation. *Journal of Consulting and Clinical Psychology, 45,* 450–454.

Tolor, A. (1956). A comparison of the Bender Gestalt Test and the digit-span test as measures of recall. *Journal of Consulting Psychology, 20,* 305–309.

Tolor, A., & Brannigan, G. C. (1980). *Research and clinical applications of the Bender-Gestalt Test.* Springfield, IL: Charles C Thomas.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9,* 260–268.

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. M. (1996). Mini-mental state examination (MMSE) and the Modified MMSE (3MS): Psychometric comparison and normative data. *Psychological Assessment, 8,* 48–59.

Townsend, J. K. (1967). The relation between Rorschach signs of aggression and behavioral aggression in emotionally disturbed boys. *Journal of Projective Techniques and Personality Assessment, 31,* 13–21.

Tranel, D. (1994). The release of psychological data to nonexperts: Ethical and legal considerations. *Professional Psychology: Research and Practice, 25,* 33–38.

Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy.* New York: Aldine.

Tryon, W. E. (1998). Behavioral assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 79–103). Boston: Allyn & Bacon.

Tucker, R. K., Weaver, R. L., Duran, R. L., & Redden, E. M. (1983). Criterion-related validity of three measures of assertiveness. *Psychological Record, 33,* 361–370.

Turk, D., & Salovey, P. (1985). Cognitive structures, cognitive processes, and cognitive behavior modification. *Cognitive Therapy and Research, 9,* 19–33.

Turkat, I. D. (1990). *The personality disorders: A psychological approach to clinical management.* New York: Pergamon Press.

Turner, S. M., DeMers, S. T., Fox, H. R., & Reed, G. M. (2001). APA's guidelines for test user qualifications. *American Psychologist, 56,* 1099–1113.

Turpin, G. (1991). The psychophysiological assessment of anxiety disorders: Three-systems measurement and beyond. *Psychological Assessment, 3,* 366–375.

Twentyman, C. T., & McFall, R. M. (1975). Behavioral training of social skills in shy males. *Journal of Consulting and Clinical Psychology, 53,* 393–401.

Ullman, L. P., & Krasner, L. A. (1965). *Case studies in behavior modification.* New York: Holt, Rinehart and Winston.

Ulrich, L. P., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63,* 100–116.

Vakil, E., & Blachstein, H. (1993). Rey Auditory Verbal Learning Test: Structure analysis. *Journal of Clinical Psychology, 49,* 883–890.

Vance, B., Fuller, G. B., & Lester, M. L. (1986). A comparison of the Minnesota Perceptual Diagnostic Test Revised and the Bender-Gestalt. *Journal of Learning Disabilities, 19,* 211–214.

Vanderploey, R. D. (Ed.). (2000). *Clinicians guide to neuropsychological assessment* (2nd ed.). Mahwah, NJ: Erlbaum.

Vanderploeg, R. D., Schinka, J. A., & Axelrod, B. N. (1996). Estimation of WAIS-R premorbid intelligence: Current ability and domographic data used in a best-performance fashion. *Psychological Assessment, 8,* 404–411.

Vanderploeg, R. D., Schinka, J. A., Baum, K. M., Tremont, G., & Mittenberg, W. (1998). WISC-III premorbid prediction strategies: Demographic and best performance approaches. *Psychological Assessment, 10,* 277–284.

Vandiver, T., & Sheer, K. J. (1991). Temporal stability of the diagnostic interview schedule. *Psychological Assessment, 3,* 277–281.

Vane, J. R., & Guarnaccia, V. J. (1989). Personality theory and personality assessment measures: How helpful to the clinician? *Journal of Clinical Psychology, 45,* 5–19.

Van Hutton, V. (1990). Test review: The California Psychological Inventory. *Journal of Consulting and Development, 69,* 75–77.

Van Zomeren, A. H., & Brouwer, W. H. (1990). Attentional deficits after closed head injury. In B. G. Deelman, R. J. Saan, & A. H. Van Zomeren (Eds.), *Traumatic brain injury: Clinical, social, and rehabilitation aspects.* Amsterdam: Swets & Zeitlinger.

Vassend, O., & Skrondal, A. (1999). The problem of structural indeterminancy in multidimensional symptom report instruments: The case of the SCL-90-R. *Behaviour Research and Therapy, 37,* 685–701.

Veiel, H. O. F., & Koopman, R. F. (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment, 13,* 356–368.

Veldman, D. J., & Kelly, S. J. (1965). Personality correlates of a composite criterion of teaching effectiveness. *Alberta Journal of Educational Research, 11,* 702–707.

Vernon, P. E. (1964). *Personality assessment: A critical survey.* London: Methuen.

Vieta, E., Colum, F., Martinez-Aran, A., Benabarre, A., Reinares, M., & Gasto, C. (2000). Bipolar II disorder and comorbidity. *Comprehensive Psychiatry, 41,* 339–343.

Viglione, D. J., & Hilsenroth, M. J. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment, 13,* 452–471.

Vilkki, J., Ahola, K., Holst, P., Ohman, J., Servo, A., & Heiskanen, O. (1994). Prediction of psychosocial recovery after head injury with cognitive tests and neurobehavioral ratings. *Journal of Clinical and Experimental Neuropsychology, 16,* 325–338.

Vincent, K. R. (1991). Black/white differences: Does age make the difference? *Journal of Clinical Psychology, 47,* 266–270.

Vincent, K. R., Castillo, I., Hauser, R., Stuart, H. J., Zapata, J. A., Cohn, C. K., et al. (1983). MMPI code types and *DSM-III* diagnosis. *Journal of Clinical Psychology, 39,* 829–842.

Vincent, K. R., & Harman, M. J. (1991). The Exner Rorschach: An analysis of its clinical validity. *Journal of Clinical Psychology, 47,* 596–599.

Visser, R. S. H. (1992). *Manual for the Complex Figure Test.* Amsterdam: Swets & Zeitlinger.

Vogt, A. T., & Heaton, R. L. (1977). Comparison of WAIS indices of cerebral dysfunction. *Perceptual and Motor Skills, 45,* 607–615.

Wagner, E. E. (1961). The interaction of aggressive movement responses and anatomy responses on the Rorschach in producing anxiety. *Journal of Projective Techniques, 25,* 212–215.

Wagner, E. E. (1971). Structural analysis: A theory of personality based on projective techniques. *Journal of Personality Assessment, 37,* 5–15.

Wagner, E. E. (1973). Diagnosis of conversion hysteria: An interpretation based on structural analysis. *Journal of Personality Assessment, 37,* 5–15.

Wagner, R. K. (1949). The employment interview: A critical review. *Personnel Psychology, 2,* 17–46.

Wagner, R. K. (1997). Intelligence, training, and employment. *American Psychologist, 52,* 1059–1069.

Wakefield, J. F. (1986). Creativity and the TAT blank card. *Journal of Creative Behavior, 7,* 127–133.

Walker, C. E., & Roberts, M. C. (2001). *Handbook of clinical child psychology* (3rd ed.). New York: Wiley.

Wallbrown, F. H., & Jones, J. A. (1992). Reevaluating the factor structure of the revised California Psychological Inventory. *Educational and Psychological Measurement, 52,* 379–386.

Walsh, K. (1994). Neuropsychological assessment of patients with memory disorders. In S. Toyz, D. Byrne, & A. Gilandas (Eds.), *Neuropsychology in clinical practice* (pp. 107–127). New York: Academic Press.

Wampold, B. E. (2000). *The great psychotherapy debate: Models, methods, and findings.* Hillsdale, NJ: Erlbaum.

Ward, L. C. (1991). A comparison of the MMPI and the MMPI-2. *Psychological Assessment, 3,* 688–690.

Ward, L. C., Ryan, J. J., & Axelrod, B. N. (2000). Confirmatory factor analysis of the WAIS-III standarization data. *Psychological Assessment, 12*(3), 341–345.

Ward, S. B., Ward, T. J., Hatt, C. V., Young, D. L., & Mollner, N. R. (1995). The incidence and utility of the ACID, ACIDS, and SCAD profiles in a referred population. *Psychology in the Schools, 32,* 267–276.

Warrington, E. K., James, M., & Maciejewski, C. (1986). The WAIS as a lateralizing and localizing diagnostic instrument. *Neuropsychologia, 24,* 223–239.

Watkins, C. E. (1991). What have surveys taught us about the teaching and practice of psychological assessment. *Journal of Personality Assessment, 56,* 426–437.

Watkins, C. E., Campbell, V. L., & McGregor, P. (1990). What types of psychological tests do behavioral (and other) counseling psychologists use? *Behavior Therapist, 13,* 115–117.

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26,* 54–60.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology, 33,* 448–457.

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1966). *Pragmatics of human communication.* New York: Norton.

Waugh, K. W., & Bush, W. J. (1971). *Diagnosing learning disorders.* Columbus, OH: Merrill.

Webb, J. T., McNamara, K. M., & Rodgers, D. A. (1981). *Configural interpretation of the MMPI and CPI.* Columbus: Ohio Psychology Publishing.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87–95.

Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children.* New York: Psychological Corporation.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary School of Intelligence.* New York: Psychological Corporation.

Wechsler, D. (1974). *Wechsler Memory Scale manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Sale of Intelligence-Revised (WPPSI-R).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children* (3rd ed.). New York: Psychological Corporation.

Wechsler, D. (1992). *Wechsler Individual Achievement Test.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997a). *WAIS-III Administration and Scoring manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale: Administration and Scoring manual* (3rd ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2001). *Wechsler Test of Adult Reading.* San Antonio, TX: Psychological Corporation.

Wedding, D. (1979). *A comparison of statistical, actuarial, and clinical models used in predicting presence, lateralization, and type of brain damage in humans.* Unpublished doctoral dissertation, University of Hawaii, Honolulu.

Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4,* 233–265.

Weed, L. L. (1968). Medical records that guide and teach. *New England Journal of Medicine, 278,* 593–600.

Weed, N. C., Butcher, J. N., McKenna, T., & Ben-Porath, Y. S. (1992). New measures for assessing alcohol and drug abuse with the MMPI-2. The APS and AAS. *Journal of Personality Assessment, 58,* 389–404.

Weed, N. C., Butcher, J. N., & Williams, C. L. (1994). Development of MMPI-A Alcohol/Drug problem scales. *Journal of Studies on Alcohol, 55,* 296–302.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist, 44,* 98–104.

Weiner, I. B. (1961). Cross-validation of a Rorschach checklist associated with suicidal tendencies. *Journal of Consulting Psychology, 25,* 312–315.

Weiner, I. B. (1966). *Psychodiagnosis in schizophrenia.* New York: Wiley.

Weiner, I. B. (1986). Conceptual and empirical perspectives on the Rorschach assessment of psychopathology. *Journal of Personality Assessment, 50,* 472–479.

Weiner, I. B. (1994). The Rorschach Inkblot Method (RIM) is not a test: Implications for theory and practice. *Journal of Personality Assessment, 62,* 498–504.

Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment, 8,* 206–213.

Weiner, I. B. (1999). Incremental validity of the Rorschach. *Assessment, 6,* 327–338.

Weiner, I. B., & Exner, J. E. (1991). Rorschach changes in long-term and short-term psychotherapy. *Journal of Personality Assessment, 56,* 453–465.

Weins, A. N., McMinn, M. R., & Crossen, J. R. (1988). Rey Auditory-Verbal Learning Test: Development of norms for healthy young adults. *Clinical Neuropsychologist, 2,* 67–87.

Weiss, L., & Price, L. (2002). *An update on the factor structure of the WMS-III.* www.psychcorp.com/catg/Wms/wmsfactor.html.

Weissman, A., & Beck, A. T. (1978, November). *Development and validation of the Dysfunctional Attitude Scale (DAS).* Paper presented at the 12th annual meeting of the Association for the Advancement of Behavior Therapy, Chicago.

Weller, R. A., Penick, E. C., Powell, B. J., Othmer, E., Rice, A. S., & Kent, T. A. (1985). Agreement between two structured psychiatric diagnostic interviews: DIS and the PDI. *Comprehensive Psychiatry, 26,* 157–163.

Wernick, R. (1955, September 12). The modern-style mind reader. *Life,* 95–108.

Weston, D. (1995). *Revision of Social Cognition and Object Relations Scale: Q sort for projective stories (SCORS-Q).* Unpublished manuscript, Cambridge Hospital and Harvard Medical School, Department of Psychiatry, Cambridge, Massachusetts.

Weston, D., & Shedler, J. (1999a). Revising and assessing Axis II. Part I: Developing a clinical and empirically valid assessment method. *American Journal of Psychiatry, 156,* 258–272.

Weston, D., & Shedler, J. (1999b). Revising and assessing Axis II. Part II: Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry, 156,* 273–285.

Wetter, M. W., Baer, R. A., Berry, D. T., Robinson, L. H., & Sumpter, J. (1993). MMPI-2 profiles of motivated fakers given specific symptom information: A comparison to matched patients. *Psychological Assessment, 5,* 317–323.

Wetter, M. W., & Deitsch, S. E. (1996). Faking specific disorders and temporal response consistency on the MMPI-2. *Psychological Assessment, 8,* 39–47.

Wetzler, S. (1990). The Millon Clinical Multiaxial Inventory (MCMI): A review. *Journal of Personality Assessment, 55,* 445–464.

Wetzler, S., & Marlowe, D. (1990). "Faking bad" on the MMPI, MMPI-2 and Millon-II. *Psychological Reports, 67,* 1117–1118.

Wetzler, S., & Marlowe, D. (1992). What they don't tell you in the test manual: A response to Millon. *Journal of Counseling and Development, 70,* 427–428.

White, G. W., Nielsen, L., & Prus, J. S. (1984). Head start teacher and aide preferences for degree of specificity in written psychological recommendations. *Professional Psychology: Research and Practice, 15,* 785–790.

White, S., & Edelstein, B. (1991). Behavioral assessment and investigatory interviewing. *Behavioral Assessment, 13,* 245–264.

Whitelaw, S., Baldwin, S., Bunton, R., & Flynn, D. (2000). The status of evidence and outcomes in stages of change research. *Health Education Research, 15,* 707–718.

Wickramasekera, I. E. (1995a). A model of people at high risk to develop chronic stress-related somatic symptoms: Some predictions. *Professional Psychology: Research and Practice, 17,* 437–447.

Wickramasekera, I. E. (1995b). Somatization: Concepts, data, and predictions from the high risk model of threat perception. *Journal of Nervous and Mental Diseases, 186,* 15–23.

Widiger, T. A., & Clark, L. A. (2000). Toward *DSM-V* and the classification of psychopathology. *Psychological Bulletin, 126,* 946–963.

Widiger, T. A., & Kelso, K. (1983). Psychodiagnosis of Axis II. *Clinical Psychology Review, 3,* 491–510.

Widiger, T. A., Williams, J., Spitzer, R., & Francis, A. (1985). The MCMI as a measure of *DSM-III. Journal of Personality Assessment, 49,* 366–378.

Wielkiewicz, R. M. (1990). Interpreting low scores on the WISC-R Third Factor: It's more than distractibility. *Psychological Assessment, 2,* 91–97.

Wiener-Levy, D., & Exner, J. E. (1981). The Rorschach Comprehensive System: An overview. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 5). San Francisco: Jossey-Bass.

Wiens, A. N. (1976). The assessment interview. In I. B. Weiner (Ed.), *Clinical methods in psychology* (pp. 3–60). New York: Wiley.

Wiesner, W., & Cronshaw, S. (1988). A meta-analytic investigation of the impact of interview format and the degree of structure on the validity of the employment interview. *Journal of Occupational Psychology, 67,* 189–205.

Wiggins, J. S. (1966). Substantive dimensions of self-report in the MMPI item pool. *Psychological Monographs, 80*(630).

Wiggins, J. S. (1971). *Content scales: Basic data for scoring and interpretation.* Unpublished raw data.

Wilkinson, G. S. (1993). *Wide Range Achievement Test-III.* Wilmington, DE: Jastak.

Williams, C. L., & Butcher, J. N. (1989a). An MMPI study of adolescents. I: Empirical validity of the standard scales. *Psychological Assessment, 1,* 251–259.

Williams, C. L., & Butcher, J. N. (1989b). An MMPI study of adolescents. II: Verification and limitations of code type classifications. *Psychological Assessment, 1,* 260–265.

Williams, C. L., Butcher, J. N., Ben-Porath, Y. S., & Graham, J. R. (1992). *MMPI-A content scales: Assessing psychopathology in adolescents.* Minneapolis: University of Minnesota Press.

Williams, J. B. W., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., et al. (1992). The Structured Clinical Interview for *DSM-III-R* (SCID). II: Multisite test-retest reliability. *Archives of General Psychiatry, 49,* 630–636.

Williams, J. M. (1991). *Memory Assessment Scales.* Odessa, FL: Psychological Assessment Resources.

Williams, R. L. (1974). Scientific racism, and I.Q: The silent mugging of the black community. *Psychology Today, 7,* 32–41.

Williamson, D. A., Veron-Guidry, S., & Kiper, K. (1998). Assessment of health-related disorders. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 256–270). Boston: Allyn & Bacon.

Wilson, B. A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. J. (1999). *Behavioural assessment of the dysexecutive syndrome manual.* Odessa, FL: Psychological Assessment Resources.

Wilson, B. A., Cockburn, J., & Baddeley, A. (1985). *The Rivermead Behavioral Memory Test.* Gaylord, MI: National Rehabilitation Services.

Wilson, T. E., & Evans, I. M. (1983). The reliability of target behavior selection in behavioral assessment. *Behavioral Assessment, 5,* 33–54.

Wing, J. K., Cooper, J. E., & Sartorius, N. (1974). *Description and classification of psychiatric symptoms.* Cambridge, England: Cambridge University Press.

Winter, D. G. (1999). Linking personality and "scientific" psychology: The development of empirically derived Thematic Apperception Test measures. In L. Gieser & M. I. Stein (Eds.), *Evocative images: The Thematic Apperception Test and the art of projection* (pp. 177–190). Washington, DC: American Psychological Association.

Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat, P. D. (1977). *Multidimensional description of child personality: A manual for the Personality Inventory for Children.* Los Angeles: Western Psychological Services.

Witmer, J. M., Bornstein, A. V., & Dunham, R. M. (1971). The effects of verbal approval and disapproval upon the performance of third and fourth grade children of four subtests of the Wechsler Intelligence Scale for Children. *Journal of School Psychology, 9,* 347–356.

Witt, J. C., & Elliott, S. N. (1983). Assessment in behavioral consultation: The initial interview. *School Psychology Review, 12,* 42–49.

Witteman, C., & Koele, P. (1999). Explaining treatment decisions. *Psychotherapy Research, 9,* 100–114.

Wolber, G. J., & Carne, W. F. (1993). *Writing psychological reports: A guide for clinicians.* Sarasota, FL: Professional Resource Press.

Wolfson, D. (1985). *Neuropsychological History Questionnaire.* Tucson, AZ: Reitan Neuropsychology Laboratory.

Wolk, R., & Wolk, R. B. (1971). *Manual: Gerontological Apperception Test.* New York: Behavioral Publications.

Wolpe, J., & Lang, P. J. (1964). A fear survey schedule for use in behavior therapy. *Behavior Research and Therapy, 2,* 27–30.

Wolpe, J., & Lang, P. J. (1969). *Fear Survey Schedule.* San Diego, CA: Educational and Industrial Testing Service.

Wolpe, J., & Lang, P. J. (1977). *Manual for the Fear Survey Schedule.* San Diego, CA: EdITS.

Wolpe, J., & Lazarus, A. (1966). *Behavior therapy techniques.* New York: Pergamon Press.

Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach Inkblot Test: A case of overstatement? *Assessment, 6,* 341–349.

Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000). The Rorschach Test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology, 56,* 395–430.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside.

Woodruff, R. A., Goodwin, D. W., & Guze, S. B. (1974). *Psychiatric diagnosis.* New York: Oxford University Press.

Woodworth, R. S. (1918). *Personal data sheet.* Chicago: Stoelting.

Woody, R. W. (Ed.). (1980). *Encyclopedia of clinical assessment* (Vol. 1). San Francisco: Jossey-Bass.

Wooten, A. J. (1983). MMPI profiles among neuropsychology patients. *Journal of Clinical Psychology, 39,* 392–406.

World Health Organization. (1990). *The ICD-10 classification of mental and behavioural disorders: Clinical description and diagnostic guidelines.* Geneva, Switzerland: Author.

Wright, D., & DeMers, S. T. (1982). Comparison of the relationship between two measures of visual-motor coordination and academic achievement. *Psychology in the Schools, 19,* 473–477.

Wright, I. C., Rabe-Hesketh, S., Woodruff, P. W., & David, A. S. (2000). Meta-analysis of regional brain volumes in schizophrenia. *American Journal of Psychiatry, 157,* 16–25.

Wurtz, R. G., Sewell, T. E., & Manni, J. L. (1985). The relationship of estimated learning potential to performance on learning task and achievement. *Psychology in the Schools, 22,* 293–302.

Wyatt, F. (1947). The scoring and analysis of the Thematic Apperception Test. *Journal of Psychology, 24,* 319–330.

Wyndowe, J. (1987). The microcomputerized Diagnostic Interview Schedule: Clinical use in an out-patient setting. *Canadian Journal of Psychiatry, 32,* 93–99.

Wysocki, J. J., & Sweet, J. J. (1985). Identification of brain-damaged schizophrenic, and normal medical patients using a brief neuropsychological screening battery. *International Journal of Clinical Neuropsychology, 7,* 40–44.

Yudin, L. W. (1966). An abbreviated form of the WISC for use with emotionally disturbed children. *Journal of Consulting Psychology, 30,* 272–275.

Zabora, J. R., Smith-Wilson, R., Fetting, J. H., & Enterline, J. P. (1990). An efficient method for psychosocial screening of cancer patients. *Psychosomatics, 31,* 192–196.

Zedeck, S., Tziner, A., & Middlestadt, S. E. (1983). Interviewer validity and reliability: An individual analysis approach. *Personnel Psychology, 36,* 355–370.

Zilmer, E. A., Bell, J. D., Fowler, P. C., Newman, A. C., & Stutts, M. L. (1991). Wechsler Verbal-Performance I.Q. discrepancies among psychiatric inpatients: Implications for subtle neuropsychological dysfunctioning. *Archives of Clinical Neuropsychology, 6,* 61–71.

Zilmer, E. A., Waechtler, C., Harris, B., Khan, F., & Fowler, P. C. (1992). The effects of unilateral and multifocal lesions on the WAIS-R: A factor analytic study of stroke patients. *Archives of Clinical Neuropsychology, 7,* 29–40.

Zimbardo, P. G. (1977). *Shyness: What it is and what to do about it.* Reading, MA: Addison-Wesley.

Ziskin, J., & Faust, D. (1995). *Coping with psychiatric and psychological testimony* (4th ed.). Los Angeles: Law and Psychology Press.

Zlotnick, C., Shea, M. T., Pilkonis, P., Elkin, I., & Ryan, C. (1996). Gender dysfunctional attitudes, social support, life events, and depressive symptoms over naturalistic follow-up. *American Journal of Psychiatry, 153,* 1021–1027.

Zolliker, A. (1943). Schwangerschaftsdepression and Rorschach'scher formdeutversuch. *Schweiz Archeives Neurologie und Psychiatri, 53,* 62–78.

Zotter, D. L., & Crowther, J. H. (1991). The role of cognitions in bulimia nervosa. *Cognitive Therapy and Research, 15,* 413–426.

Zuckerman, E. L. (1997). *The paperwork office: The tools to make your small psychotherapy practice work ethically, legally, profitably—forms, guidelines, and resources* (5th ed.). New York: Guilford Press.

Zuckerman, E. L. (2000). *The clinician's thesaurus three: A guidebook for wording psychological reports* (5th ed.). Pittsburgh: Three Wishes Press.

Author Index

Aaroe, L. A., 110
Abrams, D. M., 443, 447, 479, 480, 481, 486, 488
Abramson, L. Y., 122
Achenbach, T. M., 45, 96
Acker, M. B., 528, 532
Ackerman, P., 153, 482, 533
Ackerman, S., 480, 632
Acklin, M. N., 412, 416
Adams, H., 106, 461
Adams, K., 521
Adams, R. L., 183
Adams, W., 557
Adcock, C. J., 485
Affleck, D. C., 456
Ahn, H., 596, 598
Aiduk, R., 225
Ainsworth, M. D., 444
Albanese, A. L., 65, 596
Albert, S., 417
Alberti, L., 583
Alcock, T., 447, 461, 462, 463
Alderman, N., 573
Alessi, G. J., 118
Alfonso, V. C., 143
Alimohamed, S., 596
Allen, J., 128
Allen, R., 449, 456
Allison, J., 443, 450, 456, 457, 462
Alloy, L., 104, 121, 122
Almagor, M., 223
Alnacs, R., 101
Alonso, R. R., 109
Alp, I. E., 414
Alterman, A. I., 100
Altman, H., 85
Ambrosini, P. J., 95, 96
Ames, M. B., 444, 447, 448, 449, 451, 455, 457, 461, 462
Amick, J., 452
Anastasi, A., 3, 11, 487
Anastopoulos, A. D., 187
Anderson, E., 598
Anderson, T., 107, 137, 414
Anderson, V., 159
Andrews, D. A., 73
Andrews, H. B., 598
Angold, A., 96
Ansley, J., 575
Antonuccio, D. O., 599, 605
Antony, M. M., 74, 77, 93, 96, 579, 599
Appelbaum, M. I., 140
Apple, C., 199
Apter, A., 96
Aram, D. M., 178, 179
Arb, J. D., 160
Arbisi, P. A., 244
Archer, R. P., 65, 219, 220, 221, 225, 227, 232, 240, 271, 304, 407, 412, 414, 416, 417, 420, 473, 474, 485, 644, 646, 647, 657
Arizmendi, T. G., 596, 601
Armbuster, G. L., 448, 457, 462
Armengol, C. G., 55, 635, 646, 647
Arnau, R. C., 589
Arnold, M. B., 453, 479, 485, 506
Arredono, R. 306
Arthur, W., 74, 75
Arvey, R. D., 74, 75
Asaad, G., 40
Ascaso, E. E., 99
Ashton, M. C., 361
Atkinson, J. W., 482
Atkinson, L., 414, 416
Atlis, M. M., 67
Axelrod, B. N., 135, 184, 194, 201, 203

Baddeley, A. D., 556
Baeckman, L., 552
Baer, J., 125
Baer, R., 240, 243, 246, 247
Bagby, R. M., 240, 246, 328
Baggaley, M., 156
Bagozzi, R. P., 378, 404
Bahn, A. K., 71
Bailey, J. M., 414
Bakker, C., 124
Bakker-Rabdau, M., 12
Baldwin, S., 618
Balgopal, R., 65
Ball, J. D., 414, 417, 474
Balla, D. A., 45
Ballenger, J., 240
Balwin, A., 140
Bamgbose, O., 6, 40
Bandura, A., 121, 125
Banks, V., 479
Bannatyne, A., 152, 165
Bansal, A., 218, 228
Barber, T., 49
Barlow, D. H., 40, 74, 92, 327, 579, 599, 621, 670
Barnett, O. W., 139, 378, 405
Bar-On, R., 140
Barona, A., 183
Barregard, L., 552
Barrera, S., 446
Barrios, B. A., 111
Barth, J. T., 520, 521
Barthlow, D. L., 223, 226, 295
Bartlett, J. A., 99
Basham, R. B., 227
Basso, M. B., 210
Bates, R., 457
Baucom, D. H., 3, 125, 362
Bauer, R. M., 8, 521
Bauman, L. J., 591
Beade, L. E., 534
Beatty , 127
Beavin, J. H., 71
Beck, A. T., 62, 101, 122, 124, 587, 588, 589
Beck, J. G., 108, 122, 591, 599
Beck, S. J., 409, 421, 443, 444, 446, 447, 450, 451, 452, 453, 454, 455, 456, 461, 462, 469
Becker, C. E., 523

Bedford, D. W., 360
Beebe, D. W., 75, 139, 167, 175
Beier, E. G., 71
Bell, J., 181
Bell, R., 159
Bellack, A. S., 111, 122, 486
Bellack, L., 106, 109, 478, 479, 480, 481, 486, 488, 491, 492, 493, 503, 506, 508, 511, 514
Belter, R. W., 530
Bender, L., 529, 540
Benjamin, R., 96
Ben-Porath, Y. S., 18, 57, 219, 223, 224, 226, 228, 229, 233, 244, 268, 270, 295, 301, 306
Benson, D. F., 558
Benton, A. L., 558, 565, 566, 567
Berchick, R. J., 101
Berg, F. W., 527, 552
Berger, P., 73
Bergin, A. E., 220
Bernard, B. A., 531, 536
Bernat, E., 416
Berren, M. R., 656
Berry, D. T., 240, 247
Bersing, S., 124
Beutler, L. E., 7, 30, 33, 58, 74, 92, 212, 596, 599, 600, 601, 607, 608, 610, 611, 613, 625, 645, 646, 647, 656, 671
Bienvenu, O. J., 95
Biever, J. L., 124
Biggler, E. D., 186, 199, 560
Bilder, R. M., 574
Billings, A. G., 606
Binder, R. L., 645
Birchler, G. R., 107
Bird, D. R., 657
Bivens, A., 318
Blachstein, H., 560
Black, F. W., 531, 536
Blakley, W. A., 143
Blanchard, E. B., 606
Blashfield, R. K., 225
Blatt, S. J., 443, 450, 596
Blatter, D. D., 199
Blau, T., 41, 656
Bleich, A., 96
Blue, S., 109
Blum, G., 479
Boerger, A. R., 301
Bogojevic, G., 584
Bohannon, W. E., 397
Boland, S., 552
Boll, T. J., 182, 240
Bolton, B., 355, 360, 363
Bonarius, H., 30
Bongar, B., 65, 74
Boone, K., 520
Bornstein, A. V., 49
Bornstein, R. A., 148, 178, 179, 199, 556, 570, 572
Bornstein, R. F., 316, 320, 323, 414
Borus, J. F., 40
Botwinick, J., 552
Bowden, S. C., 178, 202, 203, 573
Bowers, D. A., 202
Bowler, R. M., 523
Bowman, M. L., 528
Boyle, R. S., 566
Bracken, B. A., 58, 59, 61, 191
Bradbury, T. N., 125
Bradley, P., 361, 378, 404
Brady, C., 228
Brandenberg, N. A., 319
Branet, E., 523
Brannigan, G. C., 537
Brar, H. S., 457
Bratu, M. E., 324
Brawman-Mintzer, O., 101
Brayfield, A. H., 50
Breen, M. J., 532
Breit, S., 124
Breitbach, L., 124
Breitholtz, E., 104, 121, 123
Brewin, C. R., 121
Bricklin, H., 43, 456
Briskin, G. C., 529, 535
Brooks, B. L., 59
Broshek, D. R., 520
Brouwer, W.H., 549
Brown, C., 576
Brown, F., 452
Brown, G. G., 571
Brown, G. K., 62
Brown, L., 73, 75
Brown, M. C., 632
Brown, R. S., 321
Brown, T., 92, 599, 646
Brtek, M. D., 76
Brungardt, T. M., 134
Brunstien, J. C., 483
Bryant, E. L., 451
Buchanan, T., 55, 56, 66
Buchenheim, P., 73
Bucholz, K. K., 96
Budd, M. A., 40
Bufka, L. F., 579, 580
Buhler, C., 446, 452, 453, 456
Buis, T., 240
Bunton, R., 618
Burgess, P., 573
Burnett, C. K., 125
Burns, G., 482
Burt, D. B., 526, 558, 563, 657
Bush, W. J., 178
Butcher, J. N., 18, 67, 219, 220, 221, 223, 224, 226, 227, 229, 231, 232, 233, 234, 238, 239, 246, 255, 268, 270, 294, 295, 301, 304, 306, 523, 581
Butters, N., 559
Byrne, D., 404

Cacciola, J. S., 100
Cacioppo, J. T., 123
Cai, G., 97
Caldwell, A. B., 232
Calestro, K., 227
Calhoun, K. S., 106
Callaghan, G. M., 579
Callender, J. C., 74
Caltagirone, C., 559
Calvin, J., 301
Camara, W. J., 3, 6, 7, 67, 129, 197, 311, 355, 358, 407, 477, 478, 485, 517, 529, 566, 587
Campbell, J., 3, 67, 104
Campbell, W., 99
Campion, J. E., 74, 75
Cancelli, A. A., 107
Canetti, C., 582
Cannell, C. F., 75
Canter, A., 534
Caputi, P., 140
Carkhuff, R. R., 71, 83
Carlson, J., 125
Carne, W. F, 625
Carnes, G. D., 457
Carpenter, D., 378
Carrera, R., 461
Carswell, L. M., 182
Cartwright, R. D., 125
Caruso, J., 135, 154
Caruso, K., 479
Caslyn, D. A., 523
Caspi, A., 74
Cass, W. A., 456
Castatino, G., 479
Castillo, A., 201, 479
Castlebury, F. D., 296
Casullo, M. M., 479
Cattell, R. B., 154
Cauffman, E., 43
Cavalli, M., 567
Cella, D., 195
Chaistain, R. L., 130
Chakraborty, D., 240
Chambers, W. J., 95
Chan, A. Y. C., 140, 412
Chaparro, N., 99
Chasain, R., 183
Chase, T. N., 569
Chastain, G. J., 130
Chelune, G. J., 199
Chen, L., 97

Chenery, H. J., 566
Cheung, S., 218
Chiles, J. A., 40
Chisholm, S. M., 223
Cho, D. W., 85, 199
Choca, J. P., 318, 322, 324, 625
Chojnacki, J. T., 224
Christiano, B. A., 124
Chui, H. C., 549
Churchard, M., 100
Chusmir, L. H., 479
Ciarlo, J. A., 378
Ciarochi, J. V., 140
Cicchetti, D. V., 11, 45
Ciminero, A. R., 106
Claassen, C. A., 56
Clark, C., 228
Clark, C. R., 561
Clark, L. A., 92
Clark, M., 295, 297
Clarkin, I. F., 33, 65, 74, 600, 601, 602, 611, 613
Clawson, A., 541
Clemence, A.J., 480
Cliff, N., 135, 154
Cloninger, C. R., 99
Clopton, A., 232
Coan, R., 451
Cobb, J., 62
Coche, E., 484
Cockburn, J., 557
Cockell, S. R., 618
Cocking, R. R., 461
Cohen, J., 73
Cohen, J. B., 453
Cohen, N. J., 179
Colby, F., 240
Colligan, R. C., 219
Collins, D., 397
Collins, J., 378, 404
Collyer, Y. M., 453
Compton, L. B., 96
Cone, J. D., 21, 22, 105, 106, 107
Conners, C. K., 45
Cook, P. E., 60, 360, 380, 402, 579
Coolidge, F., 178
Cooney, N. L., 97, 378, 611
Cooper, H., 49, 52, 108
Cooper, W., 75
Cooper, Z., 73, 124, 453, 457, 461
Corcoran, K., 10, 62
Cordoni, B. K., 153
Corkin, S., 523
Cormier, L. S., 81
Corrigan, B., 29
Coryell, W. H., 94, 95
Costa, P. T., 221, 360
Costatino, G., 478
Costello, A. J., 98
Costello, C., 450
Costello, E., 96, 98
Costello, R., 487
Cottler, L. B., 8, 96
Couch, J. V., 239
Crago, M., 596, 599, 601
Craig, P. L., 183
Craig, R. J., 316, 317, 318, 321, 322, 323
Cramer, P., 85, 480, 481, 482, 511
Crary, W. G., 85, 86
Craske, M. G., 670
Crawford, J. I., 40, 559, 579
Crites, J. O., 400
Crockett, D., 552
Cronshaw, S., 75
Cross, T., 128
Crossen, J. R., 562
Crosson, B., 193
Croughan, J. L., 96
Crowel, E. W., 187
Crowther, J. H., 124, 223
Crumpton, E., 449
Cumming, N. A., 396, 487, 580
Cunningham, T. R., 108
Cuttler, M. J., 360
Cyr, J. J., 414, 583

Dahlstrom, W. G., 50, 58, 219, 221, 222, 223, 224, 270
Daldrup, F. J., 671
Dam, H., 414
Damasio, A. R, 567
Dana, R. H., 12, 57, 411, 446, 447, 461, 478, 480
Daniel, M. H., 193, 240, 400, 552
Danielson, K., 128
Danley, W., 59
Danton, W. G., 599
Dash, V. N., 190
Datel, J. P., 398
Davidson, H., 448, 453, 455, 456, 461, 464, 467, 468
Davis, J. R., 532
Davis, R., 313, 315, 319, 321, 324, 327
Davis, W., 9
Dawes, R. M., 29
Day, J. M., 404
Dean, R. S., 532
Deardorff, N. W., 218, 295
DeFrancesco, J. J., 361, 378
Deitsch, S. E., 240
D'Elia, L. F., 520
Del Greco, L., 124
Delis, D. C., 159, 521, 559, 563, 573
Dellapieta, L., 210
DeMers, S.T., 5, 532
DeNelsky, G. Y., 599
Deniston, N. M., 359, 360
Denman, S., 557
De-Nour, A. K., 582
De Renzi, E., 567
Derogatis, L. R., 582, 583, 584, 585, 604
DeRubeis, R.J., 605
Desmarais, S., 109
DesRosiers, G., 526, 549, 554, 563, 567
Dettmer, E. L., 645
Deutsch, G., 180, 192
Devines, N. P., 40
DeVos, G., 404
Diament, C., 108
Diamond, S. P., 527
Dicken, C. F., 381
DiClemente, C., 125, 600, 602, 647
Dies, R. R., 627
Diffenderfer, D., 484
DiGiuseppe, R., 104, 107
Diller, L., 536
Dimeck, P. T., 202
DiNardo, P. A., 92
Dinning, T., 124
Dixon, W. E., 137
Dobbins, C., 164, 182
Dolman, R., 202
Dimond, M., 583
Donna, P. A., 624
Donovan, D., 378
Dorr, D., 324
Dougherty, T. W., 74, 128
Dove, A., 58
Dowd, E. T., 612, 613, 614
Dozois, D. J., 588, 589
Drab, D. L., 618
Draguns, J. G., 456, 457
Dresselhaus, T. R., 40
Driscoll, J., 131, 139
Dritschel, B. H., 124
DuAlba, L., 223, 229
Dudek, S. Z., 446, 447, 480
Duhe, A., 125
Dulcan, M. K., 98
Dumont, F., 7
Duncan, M. K., 98
Dungan, D.S., 530
Dunham, W. G., 49
Dunn, T., 324, 609
Dupont, M., 559
Dura, J. R., 584, 587
Duran, R. C., 109
Durand, V. M., 117, 119
Durflinger, G. W., 396
Durham, R. L., 199, 296
Duszynski, R. R., 448
Duthie, B., 314
Dyck, M. J., 121
Dye, O. A., 549
Dykman, R. A., 533

Earls, R., 99
Eaton, W. W., 97, 576
Ebert, R. J., 74
Ebner, D. L., 447
Eckelman, B. L., 178, 179
Edelbrock, C. S., 96, 98
Edelstein, B., 74
Edkins, G., 6, 9, 40, 580, 601
Edwards, A., 222
Edwards, D., 224, 233, 270
Edwards, G., 530
Egeland, B. R., 50
Egelkos, S., 576
Ehler, J., 94
Eichler, R. M., 443
Eidelman, 606
Eidelsonn, R.J., 125, 606
Eismen, E. J., 3, 579
Ekstrom, D. E., 85
Elashoff, J., 49
Elia, E., 566
Elion, V. H., 60
Elkin, I., 65, 606
Ellens, B. M., 55, 67, 240
Elliot, L. L., 113, 397
Ellison, R. L., 61
Elstein, A. A., 453
Elwood, R. W., 200
Emslie, H., 573
Endicott, J., 73, 91, 93, 94
Engle, D., 671
Enterline, J. P., 584
Entwisle, D. R., 7, 483
Eppinger, M. G., 183
Epstein, J. N., 45, 77, 125
Erbaugh, J., 587
Erginel, A., 464
Erickson, R. C., 523
Ernst, D., 121
Eron, L. D., 480
Eslinger, P. J., 567
Evans, J. J., 85, 87, 113, 573
Evenson, R. C., 85
Exner, J. E., 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 419, 420, 421, 422, 423, 427, 428, 433, 435, 440, 441, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475
Eyde, L., 53
Eysenck, L. D., 362, 598

Faglioni, P., 567
Fairbank, J. A., 321
Fairburn, C. G., 92, 583
Fals-Stewart, W., 328
Fantusso, J. W., 143, 156
Farmer, R. F., 104, 105, 106, 124
Farr, S. P., 575
Farrell, A. D., 552
Faust, D., 6, 26, 28, 42, 67
Feather, N., 482
Feighner, J. P., 91
Fein, D., 159
Feingold, A., 166
Feldman, S. E., 49
Fernandez-Ballesteros, R., 104
Festinger, D. S., 317
Fetting, J. H., 584
Fiel, A., 328
Finch, A. J., 530
Fincham, F. D., 125
Finger, M. S., 67
Fink, A. D., 448
Finlayson, M. A. J., 571, 657
Finn, S. E., 55, 239, 314, 623, 629, 630, 632, 642, 646, 647
Finney, B. C., 449
First, M. B., 9, 100, 101, 106, 646
Fischoff, B., 27
Fisher, D., 65, 179, 181, 201, 202, 581, 603
Fisher, R., 559
Fisher, S., 455
Fiske, D. W., 27
Flaherty, M. R., 395
Flanagan, D., 129
Flanagan, R., 479
Flaugher, R. L., 58
Fleming, B., 125, 137
Flynn, D., 618
Flynn, P., 321, 323
Foa, E. B., 599
Fode, K. L., 49
Follete, W. C., 9, 106, 601
Folstein, M. F., 86, 97, 527
Folstein, S. E., 86
Fonda, C. P., 447, 456
Fordyce, D. J., 574
Fornal, R. E., 85
Foster, S. L., 21, 22
Fowler, P., 180, 181
Fox, H., 5, 417, 559
Fraboni, M., 221
Frances, A., 9, 321
Frank, G., 416, 459
Frank, J., 598
Franzen, M. D., 209, 240, 527, 549, 656
Fraser, A. W., 49
Fray, P. J., 67
Free, M. L., 605
Freed, E. X., 536
Freedman, D., 645
French, J., 414
Friedland, R. P., 563
Friedman, A., 95, 98, 232
Friedrich, D. S., 566
Friend, R., 124
Friman, P. C., 98
Frisby, S., 53
Frueh, B. C., 417
Fuchs, D., 49
Fuld, P. A., 173, 185
Fuller, T., 532
Fuqua, D. R., 372
Furman, J. M. R., 166
Fyer, A. J., 93

Gaffney, T., 397
Gainotti, G., 559, 563
Gale, S. D., 199
Galina, H., 448
Gallagher, R. E., 129
Gallagher-Thompson, D., 598, 599
Gallucci, N. T., 222, 223
Galvin, H., 48
Galvinm, M. D., 81
Gambrill, E. D., 124
Gansler, D. A., 656
Garamoni, G. L., 121
Garb, H. N., 7, 24, 27, 28, 29, 30, 65, 66, 223, 411, 414, 480, 486, 487
Garbin, I., 588
Gardner, H., 131, 140
Garfield, S. L., 600
Garfinkel, P. E., 124
Garner, D. M., 124
Garron, G. C., 446
Garvey, M. J., 605
Gass, C. S., 199, 574, 575, 576, 632
Gaudino, E. A., 549
Gaw, K. F., 33, 603, 608, 610, 613, 615
Geary, J., 462
Geffen, G., 561, 562
Geiselman, R., 127
Geisinger, K. F., 53
Geisler, M. W., 549
Geller, J., 618
George, L. M., 420
Georgemiller, R., 193
Geradi, M. A., 606
Gerstle, R. M., 462, 465
Getter, H., 611
Gibbon, M., 8, 100
Gibertini, M., 319
Gieser, L., 483, 477
Giffen, T., 456, 469
Gill, D., 570
Gill, H., 449
Gill, M., 443
Gilley, D. W., 526
Gillin, N., 559

Gillis, N., 328
Gillum, R., 226
Gilmore, D. C., 75
Giser, S., 486
Glass, C. R., 104, 107, 123, 124, 552, 598
Glutting, J. J., 139, 156
Gocka, E., 301
Godber, T., 159
Godfrey, H. P. D., 522, 574
Goheen, H., 614
Goldberg, E., 109, 328, 378, 395, 574
Golden, C. J., 8, 85, 175, 181, 521, 523, 547
Goldfried, M. R., 106, 110, 443, 444, 447, 450, 456, 457, 461
Goldman, R. D., 61, 446
Goldring, H., 8
Goldschmid, M. L., 361
Goldstein, G., 552, 571
Goldstein, J., 405
Golightly, C., 404
Goncalves, A. A., 324
Goodglass, H., 565
Goodstein, L. D., 24, 400, 486
Goodwin, H., 91
Gordon, N., 301
Gordon, R. A., 414, 474
Gordon, W., 576
Gorman, J. M., 599
Gorsuch, R. L., 135, 143, 593
Gottesman, C., 534
Gotts, E. E., 226, 301, 304
Gottschalk, L. A., 487
Gough, H., 3
Gough, R. A., 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 367, 368, 371, 374, 377, 378, 381, 385, 386, 387, 389, 394, 395, 396, 398, 400, 404, 405, 406
Gould, R. A., 599
Gouvier, W. D., 562
Gouze, M., 226
Gow, C. A., 450
Grab, R., 6
Grace, J., 573
Graham, J., 18, 58, 60, 65, 72, 220, 223, 225,231, 232, 235, 243 301, 305,
Graham, J. R., 219, 224, 228, 240, 268, 295, 306
Gramling, S. E., 127, 552
Grant, I., 520, 570
Graves, R. E., 182, 183, 184
Gray, E. A., 104, 372
Greaves, R. E., 420
Green, C. J., 60, 62, 228
Greenbaum, P. F., 98
Greenberg, L. S., 671
Greene, R. L., 60, 218, 226, 228, 229, 232, 233, 264, 301, 306
Greenfield, N., 484
Greenspan, S., 131, 139
Greenwald, D. F., 450, 451
Greenway, P., 187
Gregory, M., 546
Gregory, R., 129
Greif, E. B., 376
Gresham, F. M., 112
Grice, J. W., 137
Griffin, R., 225, 378, 395
Grigorenko, E. L., 61, 191
Grillo, J., 321
Grimm, T., 72
Groth-Marnat, G., 6, 7, 8, 9, 30, 40, 66, 67, 68, 73, 82, 129, 138, 152, 153, 160, 171, 173, 178, 180, 181, 188, 239, 480, 482, 548, 549, 556, 570, 580, 597, 601, 621, 625, 627
Grotzer, T. A., 140
Grove, W. M., 6, 29, 414
Guarino, T., 121
Guarnacia, G., 30
Guevremont, G. C., 104, 107, 111, 122
Guilmette, T. J., 522
Guiseppe, R. D., 479
Guze, S. B., 91
Guzman, A., 319
Gwyne, J., 49
Gynther, M. D., 60, 226, 227, 228, 255, 361, 362

Haaga, D. A., 121, 122
Haan, N., 448
Hackett, H. H., 396
Hafner, A. J., 446
Hagaman, R., 601
Hain, J. D., 532, 537
Hale, J. B., 129
Haley, E. M., 456
Hall, C., 67
Hall, C. R., 202
Hall, F. D., 398
Hall, G., 218
Hall, H., 573
Hall, S., 560
Hall, W. B., 228, 229, 396
Haller, D. L., 219
Hallmark, D., 3
Hamberger, C., 378, 405
Hammeke, T. A., 8, 159
Hammen, C. L., 122
Hamsher, H., 565, 566
Handel, R. W., 57, 218, 226
Handelsman, M. M., 45, 81
Handler, L., 408, 411, 414, 418
Hansell, A. G., 452, 468
Hanson, R. K., 414
Hargrave, G. G., 224, 397
Harman, R. J., 414, 459
Harrington, S., 210
Harris, B., 180
Harris, J., 560
Harris, R., 263, 301
Harrison, P. L., 484, 485, 559
Harrison, P. L., 109
Harrower, M., 479
Hart, R., 567, 569
Harter, D. K., 104
Hartig, L., 61
Hartlage, L., 58
Hartshorne, H., 105
Harvey, V. S., 630, 631
Harwood, T. M., 596, 599, 600, 601, 603, 604, 608, 611, 615
Hass, H. D., 82, 362
Hastie, S., 26, 27
Hatch, S., 40
Hathaway, S. R., 220
Hattem, D., 116
Haverkamp, B. E., 27
Hawkins, A., 199
Hawkins, J., 179
Hawkins, K., 135, 181, 201, 202, 205, 207
Hawkins, S. A., 26, 27
Hayes, S. C., 9, 106
Haynes, J. P., 136
Haynes, S., 11, 17, 104, 106, 107, 108, 127, 487, 488
Heaton, R. K., 178, 520, 534, 555, 570, 574
Hebben, N., 644
Heckrath, L., 583
Hedlund, J. V., 85
Heflinger, C. A., 60
Heilbrun, A. B., 41, 42, 400, 401
Heimberg, R. G., 109, 123, 124, 599
Heisser, L. G., 645
Hellkamp, D. T., 532
Helmes, E., 218, 221, 556, 559
Helzer, J. E., 8, 77, 96, 97, 99
Hendin, H., 82
Henry, W., 486
Henry, B., 74, 75
Henry, E., 446
Herjanic, B., 99
Herman, P. S., 193, 483
Hersen, M., 73, 106, 109, 111, 112, 122, 420, 591
Hertz, M. R., 448, 451, 454
Herzog, D. B., 93
Hetherington, C. R., 450
Hiatt, D., 224, 397
Hibbard, M. R., 576

Hiers, T., 42
Higgins, R. L., 109
Higgins-Lee, C., 359, 402
Hildebrand, D. K., 135
Hill, A., 388
Hill, R., 396
Hillix, W. A., 50
Hillman, L., 467
Hilsabeck, M. J., 321
Hilsenroth, M. J., 296, 408, 411, 412, 418, 480
Himelstein, D. S., 462
Hishinuma, E. S., 175
Hodgson, M. J., 166
Hoffman, R. C., 361
Hogan, J., 19
Hogan, L., 55
Hogan, M. E., 532
Hogan, R., 19, 376, 397, 405
Hogarty, P., 140
Hoge, R. D., 73
Holland, A. L., 536, 565
Hollett, J., 73
Holliman, N. B., 403
Hollis, J. N., 122, 624
Hollon, S. D., 121, 126, 605
Holmes, C. B., 530
Holmstrom, R. W., 479
Holtzman, W. H., 415
Honaker, L. M., 224
Hope, D. A., 123, 599
Horn, J. L., 154, 158
Horowitz, M. J., 125
Horvarth, A. O., 614
Hosie, T. W., 59
Hostetler, S., 231, 268
Houts, A. C., 600, 646
Howard, R. C., 136
Howes, M. J., 40
Huffcutt, A. J., 74, 75
Hughes, C. W., 193
Humphrey, D. H., 224, 270
Hunkin, N. M., 202
Hunnicut, L. C., 139, 143
Hunsley, J., 220, 414
Hunter, E., 586
Hunter, J., 75, 129, 130
Hurwitz, T., 552
Hutt, M. L., 529, 533, 534, 535
Hyer, L., 318
Hynan, B. A., 447

Iacono, W. G., 127, 128
Imhof, G., 220, 417
Impara, J. C., 10, 62, 239
Ingersoll, J., 225
Ingram, R. E., 121, 122
Insua, A. M., 447
Ireland-Galman, M., 178
Iverson, G. L., 209, 211, 240, 244, 656
Ivnik, R. J., 561

Jackson, D., 71, 221, 222
Jackson, M., 559
Jacob, R. G., 606
Jacobs, J. A., 590
Jacobson, J., 188, 189, 227, 304
Jacobson, L., 57
Jacobson, N., 601
James, M., 569
Janowsky, J. S., 324, 559, 560
Jarvik, L. R., 173, 521
Jastak, S., 566
Jensen, A. L., 98
Jensen, A. R., 57, 58, 411, 522
Jewell, J., 225
Johansson, B., 104, 448, 534
John, R. S., 116
Johnson, A., 143
Johnson, C. W., 85, 86, 87
Johnson, D., 59
Johnson, J. A., 360
Johnson, J. H., 221
Johnson, J. W., 72
Johnson, L. C., 583, 584
Johnson, M. H., 94
Johnson, R. L., 99
Johnson, S. C., 199
Jones, C. H., 139
Jones, J., 357, 359, 362
Jones, M., 105
Jones, R., 124
Jongsma, A. E., 33, 40, 65
Jorgenson, A. E., 414, 474
Jung, K. G., 99

Kabacoff, R. I., 591
Kadden, R. M., 97, 378, 403, 611, 612
Kaemmer, B., 219
Kahn, M. W., 417
Kahn, R., 75
Kahn, T., 456, 469
Kalas, R. S., 98
Kallingal, A., 58
Kalnins, D., 484
Kalpakjian, C. Z., 201
Kamin, L. J., 57
Kamphaus, R., 6, 7, 45, 47, 480, 529
Kane, R. L., 66, 527
Kanfer, F. H., 72, 110, 112
Kabacoff, R. I., 592
Kaplan, E., 129, 131, 156, 159, 167, 173, 180, 181, 208, 521, 564, 565, 573, 623
Kaplan, H., 85, 327
Kaplan, R., 11, 109
Kareken, D. A., 27
Karon, B. P., 483, 484, 485, 488
Karp, S. A., 479
Karr, C., 224
Kaspar, J. C., 530, 546, 547
Katon, W. J., 40
Katsikitis, M., 62
Katz, T., 188
Kaufman, A., 44, 58, 59, 65, 130, 134, 137, 139, 140, 145, 146, 153, 154, 155, 156, 158, 159, 160, 161, 165, 166, 169, 172, 179, 180, 181, 187, 188, 192, 193, 200, 542
Kavanagh, D., 549, 554, 567
Kawas, C., 552
Kay, G. G., 66
Kazdin, A. E., 106
Kazoaka, K., 446
Keiller, S.,224, 232
Keiser, R. E., 488
Keller, M. B., 93,
Keller, S. E., 99,
Keller, J. W.,111
Kelley, C. K., 60, 228
Kelley, E., 27
Kelly, D., 446
Kelly, M. P., 199
Kelly, S. J., 396
Kelso, K., 314
Keltikangas-Jarvinen, L., 416
Kendall, P. C., 121, 122
Kesier, G., 493
Khan, F., 180
Kiecolt-Glaser, J. K., 584
Kilgore, W. D. S., 210
Kinder, B. N., 6, 417
King, T., 480, 482
Kinsella, G., 556
Kiper, B., 106
Kipnis, D., 378
Kirby, K. S., 317
Kirk, B. A., 396
Kirsch, I., 123
Klapow, J.C., 40
Klein, D. F., 93, 96, 418, 486, 487
Kleinmuntz, B., 27, 29, 30
Klepies, K., 645
Klinedinst, N., 45
Klinefelter, S. L., 219, 227
Klinetob, N., 228
Klinger, K., 487
Klinkenberg, W. D., 77
Klopfer, B., 409, 444, 445, 446, 447, 448, 450, 451, 452, 453, 455, 456, 461, 462, 464, 467, 468
Klopfer, W., 444, 484, 627, 630, 631, 642
Knight, R. G., 522, 574
Koele, P., 608
Konold, T. R., 139, 156
Koppitz, E. M., 530, 531, 532, 533, 534, 539, 540, 542, 543, 545, 546, 547

Korenblum, M., 100
Koss, M. P., 222, 563
Kostlan, A., 23, 223
Kozak, J., 552
Kramer, J. H., 71, 145, 173, 181, 521, 573
Krantz, S., 122
Krasner, L. A., 105
Kratochwill, T. R., 107, 113
Kravits, J., 191
Krishnamurthy, R., 221, 414, 420, 473
Kristjansson, B., 86
Krohn, E. J., 137
Kroll, L. S., 240
Krug, S. E., 60, 68, 239Kruse, G., 583
Kubany, E. S., 11
Kühn, R., 456
Kunce, J. T., 446, 447
Kurtines, W., 405
Kurtz, J., 153
Kurz, R. B., 463
Kotcher, S., 100
Kvaal, S., 625
Kwentus, J. A., 567

Lachar, D. 45, 58
Lacks, P., 530, 531, 532, 534, 535, 536, 537, 538, 548
Lafer, B., 361
Lai, Z. C., 200
Laird, N. M., 97
LaLone, L., 226
Lam, C. S., 201
Lamb, D. G., 240
Lambert, M. J., 40, 580, 598
Langens, T., 480
Langley, J., 74, 124
Lanning, K., 361, 386, 395
Lanyon, R. F., 24, 65, 486
Lapous, R., 71
Larrabee, G. J., 179
Larsen, K. H., 6, 124
La Rue, R., 173
Laufer, W. S., 404
Lavori, W., 93
Lazarus, A. A., 72, 106, 110, 124
Leahy, B. J., 201
Leary, T., 82, 484
Leavitt, F., 446
Ledbetter, M. F., 179, 199
Lee, J., 95, 199
Lees-Haley, P., 321
Lefebvre, D., 124
LeFever, D., 446, 452, 453, 456
Lefkowitz, J., 49
Leichenring, F., 598, 599
Leigh, J., 9
Leighton, A. H., 97
Lelos, D., 43
Lemsky, C. 41
Leninger, B. E., 552
Lenzenweger, M. F., 317
Leon, A. C., 226
Lerner, E. A., 452, 537
Lesiak, J., 532
Lester, M. L., 532
Leura, A. V., 420, 448
Levin, B., 462, 576
Levin, H., 527, 575
Levine, H., 38, 47
Levitt, F., 191, 226, 301, 304, 443, 456, 458, 462, 579
Lewak, R., 55, 232, 239, 633
Lewinsohn, P. M., 96
Lewis, J. F., 45, 49
Lezack, M. D., 139
Lezak, M. D., 65, 129, 156, 162, 165, 166, 173, 175, 177, 179, 181, 450, 527, 533, 547, 548, 549, 552, 554, 559, 560, 565, 566, 569, 572, 574, 575
Lichtenstein, S., 125
Light, B. H., 452
Lightfoot, S. L., 587
Lilienfield, S. O., 6, 7, 413, 419
Lillesand, D. V., 124
Lindsay, P. S., 121
Lindzey, G., 483, 484
Lingoes, M., 263, 301
Linscott, J., 104, 106
Linton, H. B., 449
Lipitz,, 175
Lipsitt, P. D., 43
Litchenberger, E. O., 59, 65, 134, 139, 140, 145, 146, 153, 154, 155, 156, 158, 160, 161, 169, 179, 180, 181, 187, 188, 200, 203, 204, 206, 209
Litchenstein, S., 27
Litt, M. D.,97, 378, 611
Little, K. B., 133
Llabre, M.M., 576
Lobello, S. G., 157
Lodge, J., 104, 126
Loehelin, J. C., 58
Loenberger, L. T., 41
Loftus, E. F., 27, 74, 121
Logerquist, S., 137
Long, C., 139, 522, 528, 529
Lopez, E. C., 132
Lopez, S. J., 143, 144
Loranger, A. W., 84, 320
Lorr, M., 318
Loughmiller, G. C., 61, 361
Love, K. E., 75
Lovibund, P. F., 592
Lovitt, R., 56
Lowery, N., 210
Lowman, R. L., 622
Lubin, B., 6, 7, 485
Luborsky, L., 598
Lucas, C. P., 98
Lundy, A., 483, 484, 489
Luria, A. R., 8

MaCartney-Filgate, M. S., 560
Maciejewski, C., 569
Mackintosh, M., 129
Maddox, T., 10, 62
Maher, M. C., 187
Maitra, A. K., 484
Majumber, A. R., 448
Makatura, T. J., 201
MaKinster, J. E., 445
Malgady, R. G., 478, 479
Malik, D. G., 92, 596, 646
Malina, A. C., 202
Mallinckrodtt, B., 606
Malloy, P. F., 573
Maloney, M., 3, 23, 25, 33, 440
Malpass, R. S., 191
Mankowski, P. G., 193
Mann, L., 449
Manuzza, S., 93
Marchese, M. C., 75
Margolin, G., 116, 561
Marinkovic, J., 584
Marks, P. S., 219, 226, 227, 232, 239, 633
Marley, M. L., 532, 536, 537
Marlowe, D. B., 317, 321, 328
Marmor, D., 179
Marra, C., 563
Marsh, B. J., 590
Martin, L. S., 199, 445, 463, 575
Marton, C., 100
Maruish, M. E., 40, 65, 220, 239, 579, 601
Masling, J., 413
Mason, B., 413, 447
Massman, P. J., 186, 559
Massons, J. M. D., 99
Masullo, C., 559
Matarazzo, J. D., 6, 71, 178, 183, 193
Mather, N., 44
Matheson, D. W., 190
Mathews, C. G., 520, 570
Mathur, S. R., 110
Mattson, R. E., 43, 210
Mauger, P. A., 220
Maurer, S. D., 74
May, D. J., 100, 105
Mayes, S. D., 187
Maziade, M., 101
McAllister, L., 356, 358, 360, 361, 362, 363, 367, 368, 393, 400, 403, 406
McBurnett, K., 167
McCall, R. B., 140
McCallum, R. S., 58, 61, 191

McCann, J. T., 318, 321
McCarthy, R. H., 17, 124
McClelland, P. C., 40, 480, 482, 487
McConnaughy, E. A., 618
McConnell, O. L., 546
McCord, I., 224
McCormick, I. A., 124
McCown, W., 448, 453, 465
McCracken, L. M., 549
McCrae, R. R., 221, 360, 362, 390
McCulley, E., 223
McDaniel, M. A., 74
McDermont, P. A., 139, 156
McDowell, I., 86, 412
McEvoy, L. F., 96
McFall, R. M., 124, 125
McFarland, R. A., 486
McFie, J., 181
McGarry, A. L., 43
McGlyn, F. D., 123
McGrath, R. E., 30, 225, 301
McGregor, P., 104
McGrew, K. S., 44
McHugh, P. R., 86
McIntosh, J. A., 530, 538
McKay, J. R., 100, 531
McKenna-Foley, J. M., 83
McKenna, 306
McKinney, B., 193
McLean, P., 58, 65, 130, 193, 605
McLeod, C. C., 40
McMann, R., 139
McMinn, M. R., 55, 56, 66, 67, 240, 562
McNulty, J. L., 223, 228, 233, 235, 306
McReynolds, P. A., 456
Meador, K. J., 199
Meagher, S., 62, 324
Meehl, S., 29
Megargee, E. I., 43, 60, 357, 358, 359, 362, 363, 367, 379, 380, 385, 387, 388, 389, 394, 400, 404, 405, 483, 645
Mehrabian, A., 80
Meller, P. J., 56
Mendelsohn, S., 43, 96, 380, 587
Mendoza, J., 625
Mercer, J. R., 45, 59, 61
Mergargee, E. I., 66
Merluzzi, T. V., 104, 107, 123, 124
Mermelstein, J. J., 536
Merwin, M., 199
Messick, S., 17, 19, 22
Metraux, R. W., 444, 457
Metz, C., 95
Meyer, G., 24, 225, 240, 412, 414, 416, 419, 474
Meyer, P., 358, 361, 363, 399
Meyer, R., 621
Meyerink, L. H., 576
Meyers, J. E., 223, 522
Miceli, G., 567, 559
Middlestadt, G., 74
Milberg, W. P., 131, 139, 180, 208, 644
Miller, A. S., 448, 598
Miller, H., 228, 320
Miller, H. K., 306
Miller, L., 180
Millis, S. R., 202, 207
Miller, S. M., 462
Millon, T., 18, 40, 45, 62, 311, 313, 314, 315, 317, 318, 319, 321, 322, 323, 324, 325, 326, 327, 328
Mills, C. J., 397
Milne, C. R., 106, 107, 187, 613
Mischel, W., 105
Mitrushina, M., 520, 560
Mittenberg, W., 159, 167, 182, 519, 562
Mizes, J. S., 124
Mizushima, K., 404
Moar, K. J., 561
Moberg, P. J., 521
Mock, T., 587
Moczynski, E., 656
Moes, E. J., 55, 74, 623, 646
Mogel, P., 194
Mohan, P. J., 536
Mohr, D. C., 605
Moleiro, C.,212, 600, 613
Molish, H. B., 447, 455, 458, 463
Monachesi, J. K., 220
Monahan, J. 43
Monk, M. A., 71
Monson, R. R., 97
Montross, J., 403
Moon, G. W., 143, 144
Mooney, K., 108
Moore, J. W., 559
Moos, R. H., 606
Moreland, K. L., 53
Moretti, R. J., 480
Morettis, E., 506
Morey, L. S., 225, 601
Morgan, C. D., 477
Morlan, R. K., 584
Morris, J. C., 159, 523, 566
Morrow, L. A., 523
Most, R. B., 53
Mostofosky, D., 40
Motowidlo, S. J., 76
Mouton, A., 318
Muchinsky, P. M., 75, 360
Mueller, H. H., 190
Mukerji, M., 463
Mulick, K., 188, 189
Mulvey, E. P., 43
Munley, P. H., 221
Murdoch, B. E., 566, 567
Murillo, L. G., 444, 445, 446, 457
Murphy, J. M., 583
Murphy, L., 62
Murray, H. A., 477, 481, 491
Murray, R. A., 139
Murstein, B. I., 461

Nagliere, J. A., 145, 155
Nathan, P. E., 3, 599
Neale, M. D., 531
Neff, D. F., 606
Neher, L. A., 50
Neisser, U., 20, 58, 129, 419
Nelson, C., 29
Nelson, G. E., 239, 633
Nelson, H., 183
Nelson, J., 110
Nelson, K. S., 361
Nelson, L., 57
Nelson-Gray, R. O., 104, 105, 106, 360
Neufeld, K., 97
Neuman, S., 443
Newcomb, F., 566
Newman, A., 181
Newman, F., 580
Newman, F. L., 378, 385
Newman, J. L., 372
Newport, K., 407, 530, 532, 538
Nezworski, M. T., 6, 7., 414
Nichols, D., 232, 240, 247
Nieberding, B., 3
Niederehe, G., 526
Nielsen, L., 647
Nilson, L., 552
Norcross, J. C., 598, 600, 618
Norton, J. C., 534, 552
Nottingham, E. J., 43
Noyes, R., 101, 321
Nrossi, M. P., 618
Nussbaum, P. D., 622

Oakland, T. D., 59
Oas, P., 546
Ober, B. A., 521, 563
O'Brien, W. H., 104, 106, 107, 108, 572
Obrzut, J., 487
O'Donnell, J. P., 153
Offord, K. P., 219
Ogard, E. M., 224
O'Hanlon, A. P., 561
Okun, A., 591
O'Leary, A., 108, 125, 599
Olin, J. T., 576
Oliver, J. M., 587

Ollendick, T. H., 124
O'Neal, J. H., 605
Ones, D. S., 67, 130
Orne, M. T., 108
Ortwein-Swoboda, G., 95
Osberg, T., 62
Osborne, D., 227
Osman, A., 588
Ost, L.-G., 104
Osterrith, P. A., 558
Othmer, E., 78, 82, 85
Othmer, S. C., 78, 82, 85
Otto, R. K., 41, 599
Owen, A. E., 590
Ownby, R. L., 29, 623, 624, 630, 631, 646, 64

Padesky, C. A., 108
Palmer, J., 463
Pancoast, D. C., 219, 227
Paniak, S., 572
Pantle, M. L., 447, 449, 468
Parker, J. D., 45, 412, 414, 416, 483, 571
Parkinson, A., 320
Parklinson, D., 523
Parks, C., 125, 126
Parks, R., 566
Parnell, T., 224
Parsons, O. A., 183
Pascal, G. R., 529
Pate, D. S., 566
Patinella, L., 143
Patterson, C. H., 80, 598
Paul, G. L., 160, 167, 210, 597, 619
Paunonen, S. V., 361
Peacock, E., 583
Pearson, J. S., 227
Pekarik, G., 584
Peltier, J. 487, 488
Pendleton, J. W., 109
Pennebaker, J. W., 480, 482
Penney, D. L., 55
Penrod, S., 27, 191
Perez, R. G., 99
Perkins, D. N., 140
Perry, J., 67, 320
Perry, W., 448
Persons, J. B., 108
Peters, J. R., 533
Peterson, C. A., 324
Peterson, L. M., 33, 65
Petoskey, M. D., 6
Petrocelli, J. V., 619
Peveler, R. C., 583
Pfaadt, N. K., 445
Pfeifer, C., 58
Pfinner, L. G., 167
Phillips, L., 447, 450, 456
Piedmont, R. L., 137, 360
Piersma, H. L., 317
Pilkonis, P. A., 606
Pilowski, I., 62
Pincus, H. A., 9
Piotrowski, C., 3, 6, 7, 66, 220, 311, 355, 445, 447, 480, 579
Piotrowski, Z., 445, 447, 448, 450, 456, 462
Plake, B. S., 10, 62, 239
Plant, W. T., 387
Plutchik, R., 96
Pogge, D. L., 301
Pollack, M. H., 576, 599
Ponsford, J. L., 548, 556
Ponterotto, J. G., 56
Pope, K., 55, 218, 224, 231
Porter 71
Powis, D., 301
Prabucki, K., 95
Prather, K., 488, 493, 605
Price, J. R., 321
Price, L., 202, 207
Price, P. B., 61
Prifitera, A., 148, 183, 193, 199
Prigatano, G. P., 527, 528, 571, 574, 575
Prochaska, J. O., 600, 602, 618, 619, 646
Pruit, S. D., 7, 40, 219
Puca, R. M., 480
Pugh, G., 397
Purisch, A. D., 8
Putzke, J. D., 240

Quarington, B., 414
Query, W. T., 397
Quillan, J., 124
Quinlan, D. M., 596

Radar, D. E., 143
Ramanaiah, N. V., 153, 359, 360
Ramos-Grenier, J., 53
Rankin, L. A., 125
Rao, S. M., 159, 569
Rapaport, C., 443, 447, 450, 456
Raphael, A. R., 546
Rasmusson, D. X., 552
Rathus, S. A., 124, 109, 124
Rawls, J. R., 455
Read, J. D., 121
Redden, E. M., 109, 218
Reddon, J. R., 570
Rees, L. M., 5, 656, 657
Reich, J., 101
Reich, W., 99, 100
Reichenber, W., 546
Reid, D. B., 199, 559
Reinecke, M. A., 139
Reitan, R. M., 8, 67, 129, 161, 162, 171, 172, 175, 177, 181, 182, 469, 520, 532, 549, 553, 555, 570, 571, 572, 576
Reller-Geary, L., 462
Retief, A., 478
Retzlaff, P. D., 203, 318, 319, 324, 325, 326, 327, 328, 329, 609
Reutzel, G., 395
Rey, A., 522, 558
Reyes, E., 124
Reynolds, C., 45, 58, 130, 187
Rhode, P., 417
Richard, M. T., 11
Richey, C. A., 124
Ricker, J. H., 202
Ridgeway, E. M., 453
Riessman, F., 462
Riggs, D. S., 599
Riley, E., 576
Rinaldo, J., 240
Riskind, J. H., 101
Rissmiller, D. J., 587
Roberts, B. W., 19
Roberts, M. L., 110
Robertson, R. O., 53
Robins, E., 91
Robins, L. N., 8, 77, 91, 96, 97
Robinson, D., 73
Robinson, J., 85
Robinson, L. H., 240
Rodell, J. L., 444, 457
Roe, Q., 457
Roger, R., 42, 77, 85, 92, 96, 100, 101, 240, 244, 317, 319, 328, 598, 656
Rogers, C., 71,
Rogler, L. H., 478, 479
Rohwer, T., 522
Roid, B., 199
Rolstein, L. D., 97
Romanczyk, A. J., 108
Rorschach, H., 416, 443
Rosa, L., 560
Rose, M. P., 123
Rosen, B. M., 71
Rosenberg, R., 562
Rosenhan, D. L., 646
Rosenhein, K., 153
Rosenthal, M., 77, 463
Rosenthal, R., 49, 52, 57, 108
Rosenzweig, S., 479
Ross, L. D., 75, 571
Rossi, J. S., 443
Rossini, E. D., 480, 506, 530, 546, 547
Roth, D. L., 193
Rothman, S., 129, 131
Rouse, S. V., 306
Roy, E. A., 202, 448
Royer, F. L., 536
Rozynko, V. V., 405
Rude, S. R., 599

Ruegg, R. G., 85, 86
Rufe, C. P., 398
Ruff, R. M., 571
Ruscio, J., 28
Rushton, J. P., 57
Russ, S., 48
Russell, E. W., 164, 175, 179, 182, 198, 521
Rutherford, M. J., 100, 110
Rutter, M., 72
Ryan, C., 523, 606
Ryan, J., 134, 135, 144, 160, 167, 191, 193, 194, 203, 210, 562
Ryan, M., 55
Ryan, R., 166
Rzechorzek, A., 569

Saccuzzo, D. P., 11, 109
Sacks, N. R., 93
Sadock, B. J., 327
Sahakian, B. J., 67
Saklosfe, D. H., 135
Salekin, R. T., 317
Sales, J., 42
Salmon, R., 453
Salovey, P., 73
Sameroff, A. J., 140
Sammons, M. T., 599, 605
Sandoval, R. T., 53, 56, 57
Sandy, L. R., 630, 631
Sanglade, A. A., 462, 464, 465, 474
Saperia, D., 549
Sapienza, M. M., 55
Sarchione, C. D., 360, 397
Sartorius, N., 73
Sarwer, D. B., 111, 113, 125
Saslow, G., 110, 112
Sattler, J. M., 44, 49, 50, 81, 116, 137, 144, 148, 155, 158, 161, 186, 192, 530, 622, 625, 637, 638
Satz, P., 178, 194, 560
Savage, G., 562
Savitz, K. L., 562, 582, 583
Sayers, S. L., 111, 113, 125
Saylor, C. F., 530
Sbordone, R. J., 139, 522, 523, 528, 572, 573, 657
Schachtel, E. G., 443, 457
Schafer, J., 443, 450
Schalock, R. L., 189, 645
Schechter, M. D., 527
Scheupbach, C. S., 523
Schinka, J. A., 84, 184, 226, 229, 525, 526
Schmalt, H. D., 480
Schmidt, F., 74, 129, 130
Schmidt, H., 447, 456
Schmidt, M., 199, 558, 559
Schmidt, N. B., 599, 605
Schmitz, N., 583
Schneidan, E., 477
Schneider, L. S., 576
Schrader, W. B., 58
Schuerger, J. M., 140
Schulman, S., 446
Schultheiss, O. C., 483
Schultz, C. B., 49, 560
Schumacher, J., 40, 67, 68, 73, 239, 448, 627
Schwab-Stone, M., 98
Schwarts, S., 24, 223, 606
Schwartz, G., 40, 220
Schwartz, R. M., 121
Scott, R. L., 223, 229
Schwartz, L., 191
Seat, P. D., 45
Sedlacek, W., 58
Seelen, J., 218
Seeley, J. R., 9, 96
Seeman, W., 219, 378
Seezer, M., 6
Seifer, R., 140
Seligman, M. E. P., 122, 598
Selz, M., 576
Semmel, A., 122
Senior, C., 9
Sewell, K. W., 59, 244, 317
Sewitch, T., 123
Shaffer, D., 98
Shalit, B., 448
Shanfield, S., 601
Shanley, L. A., 324
Shapiro, D. A., 449, 450, 599
Shapiro, D. M., 559
Shapiro, P., 27, 191
Shapiro, S., 530
Sharbrough, F. W., 559
Shatin, L., 457
Shaw, P. S., 362
Shea, M. T., 606
Shear, M. N., 101
Shedler, J., 320, 323
Sheehan, E. P., 328
Sheffield, J., 404
Sherman, E. M. S., 137, 526, 558
Sherman, R. H., 49
Sheslow, D., 557
Shimamura, A. P., 559
Shimkuras, A. M., 226
Shneidman, E. S., 223
Shondrick, D. D., 229
Shorkey, C. T., 124
Short, R. H., 190
Shrauger, J. S., 62
Sieber, K. O., 223
Siegel, A., 527
Siegel, L., 121, 186
Silber, D. E., 479
Silliti, J. A., 484
Silva, P. A., 74
Silver, E. J., 49, 96, 98, 591
Silveri, M., 559
Silverstein, A. B., 191, 536
Simon, W. E., 50
Simons, A. D., 605
Simpson, R. G., 530
Singer, J., 462, 463
Singer, P., 82
Singer, S., 523
Sitarenios, G., 45
Skoog, D. K., 404
Skrondal, A., 583
Slack, G. H., 455
Slate, R. J., 139, 143
Sledge, W. H., 101
Sletten, I. W., 85
Slick, D. J., 137
Smith, L. C., 573
Smith, A., 556
Smith, C.P., 228
Smith, D., 403
Smith, G., 17
Smith, G. T., 6
Smith, J., 447, 450
Smith, M., 124
Smith, M. A., 9
Smith, M. C., 7
Smith, M. L., 598
Smith, R. L., 220
Smith, S., 296, 473
Smith-Wilson, R., 584
Snelbaker, A. J., 139
Snibbe, J. R., 87
Snow, W. G., 25, 59, 182, 562, 597
Snow, R. E., 49
Snyder, W., 70, 621
Snyderman, M., 129, 131
Sobel, D. S., 40
Sobol, A. M., 97
Sokolove, S., 137
Sokolowski, K., 480
Sommer, D., 448
Sommers-Flanagan, D., 78, 81, 82, 85
Soref, E., 55
Southern, M. L., 387
Spangler, W. D., 483, 484
Sparroe, S. S., 45
Spellacy, F., 137
Spence, N., 62
Spicer, K. B., 571
Spiegler, M. D., 104, 107, 111, 122
Spielberger, C. D., 66, 590, 591, 604
Spisto, M. A., 187
Spitzer, R.L., 8, 73, 91, 93, 94, 100, 101, 321, 646
Spitznagel, E. L., 96
Spreen, O., 62, 553, 555, 558, 559, 566, 567

Springer, S. P., 180, 191
Squire, L. R., 559
Squires, N. K., 549
Staats, A. W., 104
Stafford, K. P., 229
Standage, K., 378, 403
Stanley, M. A., 591, 592
Stanton, H. C., 187
Starcevic, V., 584
Steadman, H. J., 43
Steer, R. A., 62, 101, 587, 589
Stefanyk, W. O., 570
Steffen, A. M., 598, 599
Stein, L. A. R., 228
Stein, L., 240, 306
Stein, M., 491
Stein, M. A., 139
Stein, M. I., 477, 483, 486
Stein, R. K., 591
Steinberg, M., 92, 100
Steiner, J. L., 101
Steisel, I. M., 449
Stejksal, I., 6, 414
Stella, S., 447
Stelmachers, Z. T., 645
Stephens, C., 530
Stern, F., 94
Sternberg, R. J., 60, 61, 129, 131, 140, 191
Stewart, L. E., 402, 403, 404, 559
Stoloff, M. E., 239
Stomp, T., 95
Storandt, M., 552, 566
Stormant, C. T., 449
Stotsky, B. A., 450
Stout, C. E., 573, 579
Strack, S., 315, 318
Straus, E., 137
Strauss, E., 548, 553, 555, 558, 559, 560, 566
Streiner, D. L., 320
Stricker, G., 443
Strobl, R., 95
Stroop, J. R., 522
Strupp, H. H., 71
Struss, E., 62
Stukenberg, K. W., 228, 584, 587
Stupp, H. S., 601
Sturgis, E. T., 127
Stuss, P. T., 450, 552, 575
Suchy, Y., 521
Suga, L., 148
Sullivan, D. S., 49
Sullivan, E. V., 49, 178, 199, 201, 203
Summerfeldt, L. J., 77, 93, 96
Sumpter, J., 240
Sundberg, N. D., 627
Susuki, L. A., 56, 60
Suttell, B.J., 529
Swart, E. C., 301
Swartz, E. C., 97
Sweet, H. A., 521, 523, 527
Swensen, W. M., 227
Swinson, R. P., 592
Sydeman, S. J., 590
Szasz, T., 52
Szekely, B. C., 606

Taft, R., 27
Talaga, M. C., 605
Talebi, H., 613
Tallent, N., 621, 623, 625, 626, 630, 640, 646
Talley, J., 559
Talley, P., 601
Tallman, K., 552
Tamkin, A.S., 446, 447
Tan, S. -Y., 584
Taylor, C. W., 61
Taylor, J., 49, 360, 361, 378, 567
Taylor, R., 60, 542, 543
Taylor, S., 605
Teal, M., 138, 178
Tebes, J. K., 101
Teevan, R. C., 484
Tellegen, A., 219, 224, 233, 270
Temp, G., 58
Teng, E. L., 549
Terrell, S. J., 49
Thackrey, M., 60
Thakler, C. D., 523
Thase, M. E., 605
Thelen, M. H., 7
Thiesen, J. W., 450
Thomas, A., 480
Thomas, C., 448, 453, 463
Thomas, E. A., 413
Thompson, A. P., 157, 570
Thompson, G., 447
Thompson, L., 570
Thorndike, R. L., 358
Tierney, M. C., 559, 562, 563
Timbrook, R. E., 58, 225, 228
Todd, A., 255
Todd, J., 178
Toepfer, S., 227
Togers, T., 78
Tolbert, H., 227
Tolor, A., 537
Tombaugh, T. N., 86, 656
Toner, B.B., 328
Tonsanger, M. E., 55, 632
Torgerson, S., 101
Townsend, J. K., 449
Tranel, D., 56
Tress, W., 583
Tripp, G., 104
Truax, C. B., 71
Trueblood, W., 199
Trumbo, D., 74, 75
Truumaa, A., 443, 456, 458, 462
Tryon, W. E., 114, 117
Tsui, E. C., 478
Tuck, R. R., 559
Tucker, R. L., 109
Tulsky, D. S., 179
Turk, D., 73
Turkat, I. D., 320
Turner, S. M., 5, 31, 53, 54, 67, 109, 606
Turpin, G., 128
Twentyman, C. T., 125
Tyrano, S., 96
Tziner, A., 74

Ullman, L. P., 105
Ulrich, L., 74, 75
Urbina, S., 3, 11, 487
Ustad, K. L., 244

Vagg, P. R., 590
Vakil, E., 560
Vance, B., 532
Van Denburg, E., 318, 322, 324
Vanderploeg, R. D., 23, 184
Vandiver, T., 97
Vane, J. R., 30
Van Gorp, W., 560
Van Hasselt, V. B., 591
Van Hutton, V., 358
Van Zomeren, A. H., 549
Vassend, O., 583
Veiel, H. O., 183, 185
Veldman, D. J., 396
Velicer, W., 618
Venugopala, E., 157
Vernon, P. E., 27
Veron-Guidry, S., 106
Verschell, M., 412
Vieta, E., 95
Viglione, D. J., 412, 462
Vilkki, J., 528
Vincent, K. R., 58, 225, 314, 414, 459
Vitale, A., 567
Vogt, A. T., 178
Von Baeyer, C., 122
Vriezen, E. R., 560

Waechtler, C., 180
Wagner, E., 457, 461
Wagner, R., 71, 74, 129
Wakefield, P. J., 7, 501
Walbrown, F. H., 357, 359, 362
Walker, R. N., 40, 101, 444, 457
Wallbrown, F., 29, 612, 614
Walsh, W. B., 175, 224, 557
Walzlawick, P., 71
Wampold, B. E., 596, 598, 600
Ward, L. C., 135, 191, 194

Ward, S. B., 3, 23, 25, 193, 224, 440, 587
Wassweman, J., 59
Watkins, C. E.,3, 5, 6, 104, 129, 355, 407, 418
Watkins, J. J., 156
Watkins, M. W., 139, 478
Watson, J. R., 124, 552, 571
Watts, S., 556
Waugh, R. W., 178, 559
Weatherhill, R., 480
Weaver, R. L., 109
Webb, J. T., 225, 403
Wechsler, D., 132, 133, 135, 143, 151, 153, 175, 178, 557, 566
Wedding, D., 26, 527, 549
Weed, L. L., 106, 227, 306
Weinberg, R. A., 129, 484
Weinberger, J., 484
Weiner, I. B., 24, 407, 408, 410, 413, 414, 440, 443, 447, 450, 452, 454, 455, 456, 458, 459, 462, 464, 465, 474, 576
Weiner-Levi, M., 465
Weins, A. N., 562
Weins, L. L., 202, 207, 405
Weissman, A., 122, 224
Weisz, A., 98
Weller, R., 97
Welsh, G. S., 221
Wenger, A., 319
Wenk, E. A., 405
Wernick, R., 378
Werth, T. R., 144, 194
Weschler, D., 136, 141, 146, 198, 199
Weston, D., 320, 323, 480, 481
Wetter, M. W., 240, 247
Wetzler, S., 3, 321, 323, 328
Whetzel, D. L., 74
White, G., 647
White, S., 74
Whitelaw, S., 618
Whiteman, V. L., 124
Wickramasekera, I. E., 40, 600, 647
Wicott, K. A., 487
Widiger, T. A., 9, 92, 314, 321
Widrow, L., 378
Wiedel, T. C., 24, 223
Wielkiewicz, R. M., 150, 181, 187, 574
Wiens, A. N., 72
Wiesner, W, 75
Wiggins, J. S., 294
Wilcox, N. S., 618
Wilkinson, G. S., 45, 566
Wilks, V., 566
Williams, J. B. W., 72, 101, 183, 221, 227, 231, 240, 255, 270, 295, 301, 304, 321, 581, 601, 603
Williams, C. L., 18, 226, 227, 255, 270, 295, 301, 304
Williams, J. B., 8, 100, 101
Williams, J. M., 27, 557
Williams, K., 124
Williams, R., 7, 58
Williamson, D. A,. 106
Wilson, B., 557, 573
Wilson, R. S., 556
Wilson, T., 113
Wing, J. R., 73
Winget, B. M., 50
Winter, D. E., 447, 482, 484
Wirt, R. D., 45
Wise, S. L., 613
Witmer, J. M., 49
Witt, A. C., 140
Witt, J. C., 113
Witteman, C., 608
Wolber, G. C., 625
Wolff, C. B., 584
Wolfson, D., 8, 67, 129, 161, 162, 171, 172, 175, 177, 181, 182, 520, 526, 549, 553, 555, 570, 571, 572
Wolk, R. 478
Wolk, R. B., 478
Wolpe, J., 12, 124
Wonderlich, S., 124
Wood, C., 6, 7, 413, 414, 419
Woodcock, R. W., 44, 154
Woodruff, R., 91
Woodward, M. J., 127, 203, 324
Woody, R. W., 5
Wright, D., 532
Wright, I., 522
Wurtz, R. G., 59
Wyatt, F., 480
Wylie, J. R., 417, 448
Wyndowe, J., 97
Wysocki, J. J., 527

Yamakawa, R., 175
Yeh, C. J., 93
Yesavage, J., 378
Yost, K., 116
Young, D. L., 360
Yudin, L. W., 194

Zabora, J. R., 584
Zald, D. H., 29
Zalewski, G., 3, 6, 7, 311, 355
Zalis, T., 448
Zaylor, C., 9
Zebb, B. J., 591, 599
Zedeck, S., 74
Zembar, M. J., 526
Zilmer, E. A., 180
Zimbardo, P. G., 125
Zimet, C. N., 450
Ziskin, J., 6, 42, 67
Zlotnick, C., 606
Zolliker, A., 457
Zonderman, A. B., 221, 552
Zorzitto, M. L., 559
Zotter, D. W., 124
Zuckerman, E. L., 48, 56, 81, 625
Zuroff, D. C., 596

Subject Index

Note: The bold numbers indicate primary information.

AAMD Adaptive Behavior Scale, 189
Accuracy of clinical judgments, 26–30, 505–506
Achenbach Child Behavior Checklist, 45
Achievement, 66, 130, 155, 385–387, 394–398
  college achievement, 395
  via conformance *vs.* independence, 398
  high school achievement, 66, 394
  vocational training, 396
ACID/SCAD profiles, 136, 153–154, 186–188
Acting out, 236, 252–255, 265–267, 277–278, 282, 290–291. *See also* Impulsivity
Adaptive Behavior Assessment System, 189
Adaptive Behavior Inventory for Children, 45
Adaptive testing, 66
Adjustment, 234, 402–403
Administration Checklist for the WISC-III, 144
Administration errors, 142–143
Affect Adjective Checklist, 590
Affective disorders. *See* Depression; Mania
Aggressiveness, indicators of. *See* Delinquency; Impulsivity
Agnosia, 568
Alcohol abuse, 237, 352. *See also* Substance abuse
Alexithymia, 416
Alternate forms reliability, 14–15
Alzheimer's disease, 185–187, 201, 531, 558, 560, 563, 567
Ambitents (Rorschach), 463, 471
American Association of Mental Retardation (AAMR), 189–190
American Board of Forensic Psychology, 42
American Educational Research Association (AERA), 48
American Psychological Association (APA), 239
Americans with Disabilities Act, 356
Antisocial personality. *See also* Acting out; Aggressiveness; Delinquency
  MCMI, 331–332, 339–341
  MMPI, 252–255, 274, 277–279, 290–291
  Rorschach, 454
Anxiety, indicators of:
  MCMI, 351
  MMPI, 237, 295
  Stait-Trait Anxiety Inventory, 590–593
  Symptom Checklist-90-R, 586
  TAT, 502, 510–511
  Wechsler, 164–165
Anxiety (Scale A; MCMI), 351
Anxiety Disorders Interview Schedule (ADIS), 92
Anxiety Scale (MMPI-2), 295
Anxiety Scale Questionnaire, 590–592
"Anxiety triad" (on Wechsler scales), 164–165
Aphasia, 185, 518–519, 565
Arithmetic subtest, 164, 568–570, 818
Army Alpha/Beta, 132
Artificial intelligence, 66
Assertiveness, 124–125
Assessment of Children, 622
Assessment interview and case history, 69–101. *See also* Interviewing
Attention and concentration, 210, 518, 548–556
Attention-deficit hyperactive disorder (ADHD), 136, 187
Attributional Styles Questionnaire, 122
Auditory Delayed Index, 207–208
Auditory Immediate Index, 207–208
Auditory memory, 198
Austin Mazes, 573
Automatic Thoughts Questionnaire, 122
AutoSCID-II, 100

Background Interference Procedure, 534
Bakker Assertiveness Inventory, 124
Bannatyne's factors, 151–153, 158, 678–679, 574
Barnum effect, 627
Base rate, 26, 314
BASIC ID, 72, 110, 600
Beck Depression Inventory (BDI-II), 62, 100, 122, 318, 527, 576, 580, 583, **587–590**
  interpretation, 589–590, 604
  reliability and validity, 588–589
Behavioral Assessment, 106–107
Behavioral assessment, **103–128**, 637
  behavioral interviewing, 111–114
  behavioral observation, 114–119
    event recording, 117–119, 127
    interval recording, 116–118

Behavioral assessment *(Continued)*
narrative recording, 116
ratings recording, 117–119
behavioral *vs.* traditional assessment, 103–105
cognitive behavioral assessment, 119–122
Cognitive Bias Questionnaire, 122
cognitive self-report inventories, 122–126
definition, 103
Fear Survey Schedule, 124
history and development,105–107
psychophysiological assessment, 127
Rathus Assertiveness Schedule, 124
recording cognitions, 109, 126–127
reliability and validity, 107–111
Behavioral Assessment: A Practical Handbook, 106, 111
Behavioral Assessment of the Dysecutive Syndrome, 573
Behavioral interviewing, 72, 82, 111–114
Behavioral observation, 86, 114–122
Behavioral observations (in report), 636–637
Behavior Assessment System for Children, 45
event recording, 117–119, 127
interval recording, 116–118
narrative recording, 116
ratings recording, 117–119
Behavior and Attitude Checklist, 637, 638
Bellak TAT and CAT Analysis Sheet, 506–511, 513
Bender Gestalt Test for Young Children, The, 530, 540
Bender memory (recall), 532–533, 653–654
Bender Visual Motor Gestalt Test, 521, **529–548**, 573
administration, 533–534
Background Interference Procedure, 534
emotional indicators, 530, 546–547
essential discriminators of brain damage, 529, 535–539
Angulation Difficulty, 538
Closure Difficulty, 538
Cohesion, 538
Collision or Collision tendency, 537
Fragmentation, 536
Impotence, 537
Motor Incoordination, 538
Overlapping difficulty, 536
Perseveration, 536
Retrogression, 536
Rotation, 534, 535–536
Simplification, 536
history and development, 529–531
interpretation guidelines (adults), 534–539
interpretation guidelines (children), 539–548
developmental maturation, 540–542
emotional indicators, 546–547
organicity, 542–545
visual-motor perception, 545–546
maturational guidelines, 697
memory administration, 532–533
reliability and validity, 531–533
scoring, 534–548
standardization, 530
visual-motor perception difficulties, 545
Benton Visual Retention Test, 558, 565
BEST-3 premorbid estimate, 184–185
Bias, 49, 56–62, 75–77, 130–131
Black Intelligence Test of Cultural Homogeneity (BITCH), 58
Blacky Pictures Test, 479
Block Design subtest, 172–173, 181, 532, 573
Blotto, 408
BNI Screen for Higher Cerebral Functions, 523
Borderline IQ classification, 143, 147
Borderline personality, 283–284, 348–349
Boston Naming Test, 566
Brain damage, indicators of, 178–185, 443, 523–527. *See also* Neuropsychological assessment
Brain-sensitive tests. *See* Neuropsychological assessment
Bricklin Perceptual Scales, 43
Brief instruments, **579–593**
Beck Depression Inventory, 62, 100, 122, 318, 527, 576, 580, 583, **587–590**
State-Trait Anxiety Inventory, 590–593
Symptom Checklist-90-R, 581–587, 604
selecting brief instruments, 580–581
Brief Symptom Inventory, 581–587, 604, 616. *See also* Symptom Checklist-90-R
Bulimia, 124, 288, 583. *See also* Eating disorders
Bulimia Test-Revised, 124
Butcher Treatment Planning Inventory, 581

California Neuropsychological Screening Battery, 523
California Psychological Inventory (CPI), **355–406**
assets and limitations, 360
configural interpretation, 393–406
achievement, 394–398
clinical assessment, 402–405
intellectual level, 393–394
leadership and managerial style, 399–402
social maturity, 405–406
fake bad/good, 364–365
folk concepts, 361
history and development, 356
individual scales, 370–391
1. Dominance (Do), 370–371
2. Capacity for Status (Cs), 371–372
3. Sociability, 372–373
4. Social Presence (Sp), 373–374
5. Self-acceptance (Sa), 374–375
6. Independence (In), 375–376
7. Empathy (Em), 376
8. Responsibility (Re), 377–378
9. Socialization (So), 378–379
10. Self-control (Sc), 379–381
11. Good Impression (Gi), 381–382
12. Communality, 382–383
13. Sense of Well Being (Wb), 383–384
14. Tolerance (To), 384–385
15. Achievement via Conformance, 385–386

16. Achievement via Independence (Ai), 386–387
17. Intellectual Efficiency (Ie), 387–388
18. Psychological Mindedness (Py), 388–389
19. Flexibility (Fx), 389–390
20. Femininity/Masculinity (F/M), 390–391
interpretation procedures, 363–368
MMPI (comparison with), 358–359
profile interpretation, 363–370, 393–406
profile validity, 364–365
regression equations, 367
reliability and validity, 359
social maturity, 405–406
special purpose scales, 391–393
Amicability, 392
Creative Temperament, 392
Law Enforcement Orientation, 392–393
Leadership Potential, 392
Managerial Potential, 391–392
Tough Mindedness, 393
Work Orientation, 392
standardization, 357
validity of profile 364–365
vector scale interpretation, 368–369
alphas, 369
betas, 369
deltas, 369
gammas, 369
vectors (factors), 368–369
California Verbal Learning Test (CVLT), 559
Cambridge Neuropsychological Test Automated Batteries (CANTAB), 67
Canter Background Interference Procedure. *See* Background Interference Procedure
CATEGO, 73
CERAD Battery, 523, 566
Cerebral impairment. *See* Neuropsychological assessment
Child custody, 43
Children's Apperception Test, 478, 503–505
administration, 490–491
descriptions of cards and typical themes, 503–505
Children's Auditory Verbal Learning Test, 559
Chitling Test, 58
Clang responses (associations), 163, 182, 294, 519
Clerical errors, 139, 143–144
Clinical *vs.* actuarial prediction, 29–31
Clinical Handbook of Psychological Disorders, 622
Clinical interview, 69–101. *See also* Interviewing and Clinical judgment
Clinical judgment, 25–31. *See also* Interviewer bias
accuracy of clinical judgments, 26–31
clinical *vs.* actuarial prediction, 29–31
data gathering and synthesis, 25–26
halo effect, 26
hindsight bias, 26
illusory correlation, 27
improving clinical judgment, 28–29
interpretation, 33–36
primacy effect, 26
self-fulfilling prophecy, 26
Clinical Neuropsychology: A Pocket Handbook for Assessment, 622
Clinical Practice of Career Assessment, The, 622
Clinician Rating Form, 581, 603
Clinician's Handbook, The, 622
Code types (MMPI-2/MMPI-A). *See* Two point code types
Coding Digit Symbol. *See* Digit Symbol
Cognitive behavioral assessment, 119–127
recording cognitions, 126–127
self-report inventories, 122–126
Cognitive Bias Questionnaire, 122
Cognitive Error Questionnaire, 124
Cognitive processes of depression, 122–123
COGSCREEN, 66
Commonsense problem solving, 140
Compendium of Neuropsychological Tests, 62
Competency, 42
Competency Screening Test, 43
Competent use of assessment instruments, 53–55
Complexity of problem (and treatment planning), 602, 608–610
Comprehension subtest, 167–168
Comprehensive Norms for an Expanded Halstead-Reitan Battery, 520
Comprehensive System. *See* Rorschach
Compulsive, 342–344, 455. *See also* Obsessive-compulsive disorder
Computer-assisted assessment, 9, **66–68**, 77, 231, 234, 358, 521, 627
Computer Use in Psychology: A Directory of Software, 239
Concentration. *See* Attention
Conceptual validity, 25
Concrete thinking, 163, 182
Concurrent validity, 19
Confirmatory bias, 75
Conflict Resolution Inventory, 124
Confrontation during interviewing, 83
Conners' Parent Rating Scale-Revised, 45
Conners' Teacher Rating Scale-Revised, 45
Construct validity, 21
Consulting Psychologists Press, 673–674
Content scales (MMPI-2/MMPI-A), 218–236, 294–300
Content validity, 18
Content variance, 222
Controlled Oral Word Association (COWA) test, 527, 566–569
administration, 567
interpretation, 568–569
norms, 568–569
reliability and validity, 567
Convergent validity, 21–22
Conversion disorder, 273
"Conversion V" (MMPI), 234, 272–273, 575
Coping style (and treatment planning), 236–237, 610–612

Cost-effectiveness (financial efficacy), 9, 40, 67, 130, 579, 580, 601
Creative Temperament (CPI), 392
Criterion contamination, 21
Criterion validity, 19
Critical items (MMPI-2), 233, 235
Crystallized intelligence. *See* Fluid *vs.* crystallized intelligence; Horn groupings

Data collection, 32
Declarative memory, 197
Defense mechanisms, 404, 511
Defense Mechanisms Manual, 511
Defined code types, 224, 235–236
Delinquency, 378, 404. *See also* Impulsiveness
Delis-Kaplan Executvie Function System, 573
Dementia. *See* Alzheimer's disease
Dementia Assessment Battery, 523
Denman Neuropsychology Memory Scale, 557
Dependent personality, 170, 334
Depression, cognitive processes of, 122–123
Depression, indicators of:
  Beck Depression Inventory, 587–590
  behavioral assessment, 122–123
  CPI, 403
  MCMI, 333
  MMPI, 248–250, 271–272, 276–283, 296, 298
  Rorschach, 414, 453, 456, 473
  Symptom Checklist-90-R, 585
  TAT, 493, 500
Depression Anxiety Stress Scale, 588
Depression and memory, 517
Descriptive statistics (Rorschach), 694–696
Developmental Bender Gestalt Test scoring system (Koppitz), 539
Developmental Test of Visual Motor Integration, 532
Deviation IQ, 141–142
Diagnostic Interview for Children and Adolescents (DICA), **99–100**
Diagnostic Interview Schedule (DIS), 77, **96–99**
Diagnostic Interview Schedule for Children (DISC), 77
Diagnostic and Statistical Manual of Mental Disorders (DSM), 73, 78, 97, 225, 238, 314–316, 319, 321, 324, 420, 581, 587, 591, 604, 609, 645, 648
Diagnostic and Structured Interviewing: A Handbook for Psychologists, 92
DIANO III, 73
Dictionary of Behavioral Assessment Techniques, 111
Differential therapeutics, 599
Digit Span subtest, 165, 181, 556, 558, 564, 568
Digit Symbol-Coding subtest, 170–172, 181, 532, 556, 558, 564, 568
Directive *vs.* nondirective interviewing, 82–83
Direct questioning, 82
Discriminate validity, 21–22
Disorders of Personality, 313, 315, 324, 327
Dissimulation. *See* Faking bad; Malingering
Dissociative disorders, 285
"Distress scales," (MMPI-2/MMPI-A), 237, 280–281
"Dodo bird" verdict, 598
Dominance, 370–371
Draw-A-Person. *See* Human Figure Drawings
DSM (and *DSM-IV*). *See Diagnostic and Statistical Manual for Mental Disorders*
Dyadic Attribution Inventory, 125
Dysfunctional Attitudes Scale, 124

Eating Attitudes Test, 124
Eating disorders, 124, 583
Eating Disorders Examination, 92
Ecological (everyday) validity, 139
Educational context, 44–46
Educational and Psychological Measurement, 10
Edwards Personal Preference Schedule, 613
Ego Strength Scale, 305
Electroencephalogram (EEG), 127
Emotional indicators (on Bender), 530, 546–547
Emotional intelligence, 140
Empathy, 376, 401
Empirical criterion keying, 215
Episodic memory, 197
Equal Employment Opportunity Commission (EEOC), p. 57
Essential discriminators of intracranial damage, 529
Estimated Learning Potentials, 59
Ethical practice of assessment, **48–56**, 67
  communicating test results, 55–56
  competencies in using, 53–54
  integration and use of results, 54–55
  invasion of privacy, 50–51
  inviolacy, 51–52
  labeling and restriction of freedom, 52–53
  professional relationship, 48–50
  test security, 56
Ethical Principles of Psychologists and Code of Conduct, 48, 51
Ethnicity and test performance, 50, 56–62, 131–132, 139, 191–192, 218, 226–229, 324, 358, 478
Evaluating psychological tests, 10–22
  norms, 11–12
  practical considerations, 11
  reliability, 12–17
  standardization, 11–12
  theoretical orientation, 11
  validity, 17–25
Event recording, 117
Everyday memory, 199
Executive Functions, 197, 572–574
Exner's Comprehensive System. *See* Rorschach
Expectancy effect, 49
Explicit memory, 197
Externalizing coping style, 610–612
Extratensives (Rorschach), 463, 468, 471
Extraversion-introversion, 268, 368–369, 372–373. *See also* Introversion-extraversion

Face validity, 18
Factor analysis, 21, 362

Faking bad, 240–247, 325, 329, 356, 364–365, 383, 416–417
Faking good, 240–247, 325, 329, 356, 381
Family Apperception Test, 479
Family of Man, 479
FAS test. *See* Controlled Oral Word Association test
Fb (F back) Scale, 243–244
Fear of Negative Evaluation Scale, 124
Fears and anxieties, 123–124
Fear Survey Schedule, 124
Fear Survey Schedule for Children, 124
  Revised Fear Survey Schedule for Children (FSSC-R), 124
Feedback of test results, 54–56, 632–633
Finger tapping, 527, 570–572
  administration, 571
  interpretation, 572
  norms, 572
  reliability and validity, 570–571
Five Factor model, 359–360
Fixed *vs.* flexible batteries in neuropsychology, 8, 521
F-K index, 246
Flexible battery approach, 521
Fluid *vs.* crystallized intelligence, 154, 678–679
Folk concepts (CPI), 361
Fp (Infrequency-Psychopathology) Scale, 244
Frontal lobe damage, 573
Frontal Lobe Personality Scale, 573, 575
Frostig Developmental Test of Visual Perception, 532
F Scale (Infrequency), 242–243
Fuld profile, 185
Future of assessment, 9

Galvanic Skin Response (GSR), 127
Gambrill Assertion Inventory, 124
Gender-Role Scales (GM/GF), 307
General Abilities Measure for Adults, 58
General Health Questionnaire, 583
General medical setting, 39–41
General Memory Index, 204, 206, 209
Geriatric Depression Inventory, 576
Gerontological Apperception Test, 478
Gifted children, 190
Global Severity Index, 582, 584, 616. *See also* Symptom Checklist-90-R
Guidelines for Computer-Based Test Interpretations, 48, 68

Halo effects, 75
Halstead Reitan Neuropsychological Test Battery, 8, 520, 523
Hamilton Rating Scale for Depression, 576, 588
Handbook of Behavioral Assessment, 106, 520
Handbook of Clinical Child Psychology, 622
Harris and Lingoes subscales, 218, 222, 232, 236, 300–304
Healy Picture Completion, 132
High Risk Model, 600–601
Hindsight bias, 27
Histrionic personality, 164, 250–252, 283–284, 336–338
"Hold" *vs.* "no-hold" tests, 166, 178
Holtzman Inkblot Test, 415
Homosexuality, 286, 501
Hopkins Symptom Checklist, 581
Horn groupings, 154–155, 678–679
Hostility. *See* Aggression
House Tree Person, 529, 548
Human Figure Drawings, 527, 529, 573
Hypnosis, 498
Hypochondriasis, 247–248, 271, 273–276
Hysteria, 250–252

ICD-10, 645
Illness Behavior Questionnaire, 62
Imagery (clinical assessment of), 125
Immediate Memory Index, 206, 208
Impairment index, 8
Impulsivity, indicators of:
  Bender, 530
  CPI, 378–381
  MMPI, 252–255, 265–267, 277–278, 282, 290–291
  Rorschach, 462
Incremental validity, 23–25
Individual Neuropsychological Testing for Neurotoxicity Battery, 523
Information subtest, 166–167, 182, 558
Information variance, 91
Insanity, 41–43
Insight and judgment, 89
Intellectual disability, 188–190. *See also* Mental retardation
Intelligence, 140–143
  bias, 56–62, 130–132
  classifications, 141–143
  deviation IQ, 141–142
  ethnicity, 52–56, 130–132
  fluid *vs.* crystallized intelligence, 154, 678–679
  meaning of IQ scores, 141–142
  misuse, 131
  pros and cons of testing intelligence, 129
Intelligence Quotient (IQ), 141–142
Internalizing coping style, 610–612
International Adaptations of the MMMPI-2, 229
Interpersonal relations, 237, 438–439
Interscorer reliability, 15–16
Interval recording, 116–118
Interviewing, **69–101**, 111–114, 523–527
  assets and limitations, 76–78
  behavioral interviewing, 111–114
  checklist for assessment interview and case history, 79
  concluding, 84
  confrontation, 83
  directive *vs.* non-directive, 80, 82
  history, 78–81
  interpreting interview data, 90–91
  interview tactics, 81–84
  mental status exam, 84–90

Interviewing *(Continued)*
neuropsychological interviewing, 523–527
note taking, 80–81
open-ended questions, 82
preliminaries, 81–82
reliability and validity, 74–76
structured clinical interviews, 76, 78–80, **91–101**
Introversion-extraversion, 268, 368–369, 372–373
Introversives (Rorschach), 463, 468, 471
Invasion of privacy, 50–51
Inviolacy, 51–52
IQ classifications, 142–143
IQ-memory differences, 204
Irrational Beliefs Test, 124

Journal of Behavioral Assessment, 106–107
Journal of Behavior Therapy and Experimental Psychiatry, 106
Journal of Clinical Psychology, 480
Journal of Personality Assessment, 10, 48, 412
Journal of Psychoeducational Assessment, 10

K (Correction) Scale, 245–246
Kaplan & Sadock's Pocket Handbook of Clinical Psychiatry, 327, 622
Kaufman Assessment Battery for Children (K-ABC), 44, 136, 189
Kaufman Test of Educational Achievement, 189
Koh's Block Design, 132
Koppitz scoring system (Bender), 531–532, 539
Korsakoff's syndrome, 201
Kuder-Richardson reliability, 13

L (Lie) Scale, 244
Labeling and restriction of freedom, 52–53
Law Enforcement Orientation, 392–393
Lawrence Psychological Forensic Examination, 84
Law School Aptitude Test, 61
Leadership, 399–402
Leadership Potential, 392
Learning and memory, 197, 210
Learning disabilities, 151–154, 165, 186–188
Legal context, 41–44, 648, 655–657
Leiter International Performance Scale, 58
Letter-Number Sequencing subtest, 168, 211, 558, 564, 568
Logical Memory subtest (WMS-III), 211
Luria Nebraska Neuropsychological Test Battery, 8, 521
Luria's theories of intelligence, 140

Malingering, 210, 242–244, 273, 417
faked memory loss, 210
Managed care, 579, 601
Managerial Potential, 391
Mania, indicators of:
MCMI, 351
MMPI, 265
Manifest Anxiety Scale, 590–592
Manual for Using the MMPI-2 as a Therapeutic Intervention, 633
Marital assessment, 125
Marital Attitude Survey, 125
Marital relationships, 125
Mastery Model, 144
Matrix Reasoning subtest, 174
Mazes subtest, 177–178
McGill Pain Questionnaire, 62
McNaughton rule, 43
Measures for Clinical Practice: A Sourcebook, 10, 62
Medical School Aptitude Test, 61
Medical settings, 39–41
Memory (and learning), 197, 210, 556–565
Memory Assessment Scales, 557
Memory impairment and depression, 558, 563
Memory Quotient, 198–199
Mental Measurements Database, 62
Mental Measurements Yearbook, 10, 62, 239, 355
Mental retardation, 188–190. *See also* Intellectual disabilities
Mental status exam, 70, **84–90**
Meyer, Adolf, 85, 127
Michigan Alcoholism Screening Test, 318
Millon Adolescent Clinical Inventory, 45
Millon Behavioral Health Inventory, 40, 62
Millon Behavioral Medicine Diagnostic, 40
Millon Clinical Multiaxial Inventory, **311–353**, 607, 609, 613
assets and limitations, 319–323
Base Rate scores, 314
Clinical Personality Patterns, 329–346
Aggressive (Sadistic, Scale 6B), 341–342
Antisocial (Scale 6A), 339–341
Avoidant (Scale 2A), 331–332
Compulsive (Scale 7), 342–344
Dependent (Scale 3), 334–336
Depressive (Scale 2B), 333–334
Histrionic (Scale 4), 336–338
Narcissistic (Scale 5), 338–339
Passive-Aggressive (Negativistic, Scale 8A), 344
Schizoid (Scale 1), 329–331
Self-Defeating (Scale 8B), 345–346
Clinical Syndromes, 351–352
Alcohol Dependence (Scale B), 352
Anxiety (Scale A), 351
Bipolar: Manic (Scale N), 351–352
Drug Dependence (Scale T), 352
Dysthymia (Scale D), 352
Post-Traumatic Distress Disorder (Scale R), 352
Somatoform (Scale H), 351
history and development, 313–316
interpretation procedure, 323–328
Modifying Indices (Validity Scales), 328–329
Debasement Index (Scale Z), 329
Desirability Index (Scale Y), 329
Disclosure Index (Scale X), 329
Validity Index (Scale V), 328
noteworthy responses (critical items), 327
profile validity, 328–329
reliability and validity, 317–319
Severe Personality Pathology, 346–351

Borderline (Scale C), 348–349
Paranoid (Scale P), 349–351
Schizotypal (Scale S), 346–347
Severe Syndromes, 353
Delusional Disorder (Scale PP), 353
Major Depression (Scale CC), 353
Thought Disorder (Scale SS), 353
standardization, 315
theoretical considerations, 315
treatment planning with, 327
Mini Mental Status Examination, 86, 527
Minnesota Multiphasic Personality Inventory (MMPI/MMPI-2/MMPI-A), **213–309**, 350–357, 519, 527, 575–576, 607, 609, 611, 613, 616, 629, 637
administration, 45, 230–232
age groups, use with, 226–227
assets and limitations, 223–226
clinical scales, 214, 247–270
Scale 1: Hypochondriasis (Hy), 247–248
Scale 2: Depression (D), 248–250
Scale 3: Hysteria (Hy), 250–252
Scale 4: Psychopathic deviate (Pd), 252–255
Scale 5: Masculinity-Femininity (Mf), 255–258
Scale 6: Paranoia (Pa), 258–261
Scale 7: Psychasthenia (Pt), 261–263
Scale 8: Schizophrenia (Sc), 263–265
Scale 9: Hypomania (Ma), 265–267
Scale 0: Social Introversion (Si), 236, 268–270
computerized interpretation, 231, 234, 239–240
content scales, 218, 236, 294–300
critical items, 236, 304
Harris and Lingoes subscales, 218, 222, 232, 236, 300–304
MMPI-A content scales, 220, 297–300
MMPI-2 content scales, 295–297
Social Introversion subscales, 236
Wiggins content scales, 218
critical items, 233, 235, 304
cross-cultural/national use, 218, 226, 229
cry for help, 242–243
defined code types, 224, 235–236
ethnicity (use with minorities), 218, 227–229
faking bad/faking good, 240–247
history and development, 215
interpretation procedure, 231–239
K correction, 240
malingering, 242–244
Overcontrolled Hostility Scale (OH), 288
profile definition, 227–228
reliability and validity, 220–223
restandardization, 218–219, 224
scoring directions, 232–233, 690
short forms, 231
social class and education, 229–230
supplementary scales, 235–236, 304–309
A/Anxiety (A), 305
AAS/Addiction Acknowledgement Scale, 306
APS/Addiction Potential Scale, 306
Do/Dominance, 307
Es/Ego Strength (Es), 305
GF/Feminine Gender Role, 307
GM/Masculine Gender Role, 307
MAC-R/MacAndrew Scale, 305–306
MDS/Marital Distress Scale, 306
Mt/College Maladjustment, 307
O-H/Overcontrolled Hostility Scale, 306–307
PK/Posttraumatic Stress Disorder Scale, 307
Re/Responsibility, 307
R/Repression (R), 305
treatment implications, 238–239
two point codes, 224–225, 233, 270–294
12/21, 271–272, 575
13/31, 272–274, 575
14/41, 274
18/81, 275
19/91, 275–276
23/32, 276–277, 575
24/42, 277–278
26/62, 279
27/72, 280–281, 616
28/82, 281–283
29/92, 282–283
34/43, 283–284
36/63, 284
38/83, 285–286
45/54, 286–287
46/64, 287–288
47/74, 288
48/84, 288–289
49/94, 290–291
68/86, 291–292
69/96, 292–293
78/87, 293
89/98, 294
validity scales, 233, 240–247
Cannot Say (Cs), 240–241
Fb (and F1 and F2), 243–244
F-K index (Dissimulation Index), 246
Fp (Infrequency-Psychopathology) Scale, 244
F Scale, 242–243
K Scale, 245
L Scale, 244
S (Superlative) Scale, 246–247
True Response Inconsistency Scale (TRIN), 240–241
Variable Response Inconsistency Scale (VRIN), 241
Minnesota Multiphasic Personality Inventory (MMPI-2 and MMPI-A) Content Scales, 295–300
A-aln/Adolescent-Alienation, 298
A-ang/Adolescent-Anger, 298
A-anx/Adolescent-Anxiety, 297
A-biz/Adolescent-Bizarre Mentation, 298
A-con/Adolescent-Conduct Problems, 299
A-cyn/Adolescent Cynicism, 299
A-dep/Adolescent-Depression, 298
A-fam/Adolescent-Family Problems, 300
A-hea/Adolescent-Health, 298
A-las/Adolescent-Low Aspirations, 299

Minnesota Multiphasic Personality Inventory (MMPI-2 and MMPI-A) Content Scales *(Continued)*
A-lse/Adolescent-Low Self-Esteem, 299
ANG/Anger, 296
ANX/Anxiety, 295
A-obs/Adolescent-Obsessiveness, 297
A-Sch/Adolescent-School, 300
A-Sod/Adolescent-Social Discomfort, 299–300
ASP/Antisocial Practices, 296
A-trt/Adolescent-Negative Treatment Indicators, 300
BIZ/Bizarre mentation, 296
CYN/Cynicism, 296, 299
DEP/Depression, 296
FAM/Family Problems, 297
FRS/Fears, 295
HEA/Health Concerns, 296
LSE/Low Self Esteem, 296
OBS/Obsessiveness, 295
SOD/Social Discomfort, 297
TPA/Type A, 296
TRT/Negative Treatment Indicators, 297
WRK/Work Interference, 297
Minnesota Multiphasic Personality Inventory (MMPI-2 and MMPI-A) Supplementary Scales, 305–308
A/Anxiety (MMPI-A), 308
A/Anxiety (MMPI-2), 305
AAS/Addiction Acknowledgement Scale, 306
ACK/Alcohol Drug Acknowledgement Scale, 308
APS/Addiction Potential Scale, 306
Do/Dominance, 307
Es/Ego Strength, 305
GF/Feminine Gender Role, 307
GM/Masculine Gender Role, 307
IMM/Immaturity Scale, 308
MAC-R/MacAndrew Alcoholism Scale, 308
MAC-R/MacAndrew Alcoholism Scale-Revised (MMPI-2), 305–306
MDS/Marital Distress Scale, 306
Mt/College Maladjustment, 307
O-H/Overcontrolled Hostility, 306–307
PK/Posttraumatic Stress Disorder, 307–308
PRO/Alcohol Drug Proneness Scale, 308
PS/Posttraumatic Stress Disorder, 308
Re/Responsibility, 307
R/Repression (MMPI-A), 308
R/Repression (MMPI-2), 305
Mini-SCID, 100
Minorities, 50, 56–62, 131, 139, 190–191, 218, 226–229, 358, 478. *See also* Ethnicity; Test performance
Missouri Automated Mental Status Examination Checklist, 85
Misuse of psychological tests, 50–51, 131
MMPI-2 in Psychological Treatment, 238
Modern Psychopathology, 313
Motivation Assessment Scale (MAS), 117, 119–121
Motor abilities, 569–572
Multiple cutoffs (in test selection), 30
Multiple regression equations (in test selection), 30

Narcissistic personality, 338–339
Narrative recording, 116
National Adult Reading Test (NART), 183–184
National Computer systems, 232, 301, 303, 674
Naysaying, 222
Needs, 481
Negative predictive power, 318–319
Neglect (visual), 173
Neurobehavioral Evaluation System, 66
Neurobehavioral Rating Scale, 527, 575
Neurological impairment. *See* Neuropsychological assessment
Neuropsychological Assessment, 62
Neuropsychological assessment, 8, 40, 178–188, **517–577**
Beck Depression Inventory, 521, 527, 576
Bender Visual Motor Gestalt (Bender), 521, 529–548
BNI Screen for Higher Cerebral Functions, 523
Controlled Oral Word Association (COWA) test, 527, 566–569
domains of neuropsychological assessment functions:
emotional status, 518, 574–576
executive functions, 518, 572–574
memory and learning, 518, 556–565
mental activities (attention and speed of information processing), 518, 548–556
motor ability, 518, 570–572
verbal functions and academic skills, 518, 565–570
visuoconstructive abilities, 518, 528–548, 529
executive functions, 518
finger tapping, 527, 570–572
fixed battery approach, 521
flexible hypothesis testing approach, 8, 521
free drawing procedures, 527
Halstead Reitan Neuropsychological Test Battery, 520, 523, 549
history and development, 520–523
Human Figure Drawings, 529, 548
impairment *vs.* disability, 522
interviewing for brain impairment, 523
localization, 520
Luria Nebraska Neuropsychological Battery, 521
Mini-Mental State, 527
Minnesota Multiphasic Personality Inventory (MMPI), 527, 575–576
Neurobehavioral Assessment Format, 527
Neurobehavioral Rating Scale, 527
Neuropsychological History Questionnaire, 526
Neuropsychological Status Exam, 526–527
Neuropsychological Symptom Checklist, 525, 527
pathognomonic sign approach, 521
Patient Competency Rating, 527, 574

Rey Auditory Verbal Learning Test (RAVLT), 522, 527, 558–565
Rey-Osterrith Complex Figure Test, 522, 558
screening battery (recommended), 517–518, 527–576
Stroop Test, 522
Trail Making, 520, 523, 527, 549–556
WAIS-III/WISC-III, 519, 527
Wechsler Memory Scale, 135, 138, **197–212**, 523
Neuropsychological Assessment in Clinical Practice, 517, 548, 622
Neuropsychological History Questionnaire, 526
Neuropsychological Status Exam, 84, 526–527
Neuropsychological Symptom Checklist, 525, 527
Neuropsychology Behavior and Affect Profile, 575
Neuropsychology Review, 65
Neurotic triad (MMPI-2/MMPI-A), 225, 271, 274, 293, 575
NIMH Core Neuropsychological Battery, 523
Norm-favoring *vs.* norm-questioning, 368
North Carolina Mental Status Examination, 85

Object Assembly, 176–177
Obsessive, 474. *See also* Obsessive-compulsive disorder
Obsessive-compulsive disorder, indicators of:
MCMI, 342–344
MMPI, 261–263, 280–281, 293, 295, 297
Rorschach, 474–475
Symptom Checklist-90-R, 585
Office of Federal Contract Compliance, 57
Office of Science and Technology, 50
Open-ended questioning, 82
Organic impairment. *See* Neuropsychological assessment
Orientation, 88, 209
Outcome Questionnaire-45 (OQ-45), 581
Overcontrolled Hostility Scale (OH), 288

Paper Office, The, 48, 56
Paranoia, indicators of:
MCMI, 349–351
MMPI, 258, 272–279, 287, 291–292
Rorschach, 474
Symptom Checklist-90-R, 583
Paraphasias, 565
Parkinson's disease, 201–202, 576
Passive aggressive personality, 273, 344–345
PASS model, 140
Pathognomonic sign approach, 521
Patient Competency Rating, 527
Patterns of test usage, 3, 5–10
behavioral assessment, 8–9
behavior therapy, 8
child assessment, 7
clinical interview, 7–8
neuropsychological assessment, 8
projective techniques, 7
structured interviews, 7–8
Peabody Picture Vocabulary Test-III, 58, 137
Percentile rankings for Wechsler deviation IQs, 142–143, 643, **677**
Perceptual Organization Index, 149
Performance-Verbal differences. *See* Verbal-Performance differences
Perseveration, 459, 536–537, 554, 573
Personality Disorder Examination, 81
Personality disorders *vs.* personality style, 322
Personality Inventory for Children, 45
Person perception accuracy, 27
Phases in clinical assessment, 33–36
Picture Arrangement subtest, 174–176, 181
Picture Completion subtest, 170, 182, 568
Pinther-Paterson Test, 132
Pittsburgh Occupation Exposure Test, 523
Planning-Attention-Successive-Simultaneous (PASS) model, 140
Positive predictive power, 318–319
Post Traumatic Stress Disorder Scale (PK/PS), 308
Predictive validity, 19
Premorbid level of functioning, 182–185
Present State Examination, 72
Press, 481
Primacy effect, 27
Privacy and Behavioral Research, 50
Process approach, 131, 180–181
Processing Speed Index, 150–151, 179, 181–182
Projective techniques. *See* Rorschach; Thematic Apperception Test
Psychiatric Diagnosis, 91
Psychiatric setting, 38–39
Psychodiagnostik, 408
Psychological assessment *vs.* psychometric testing, 3–5, 47
Psychological Assessment Resources, 675
Psychological clinic (context), 46–48
Psychological Corporation, 675
Psychological report, 621–671. *See also* Report writing
Psychological report writing. *See* Report writing
Psychometric testing *vs.* psychological assessment, 3–5
Psychophysiological assessment, 127–128
Psychosis. *See* Schizophrenia
Psychosomatic disorders, 403–404
Psychware Sourcebook, 68, 239
Publication Manual of the American Psychological Association, 624
Publishers (of tests), 673

Rathus Assertiveness Schedule, 124
Ratings recording, 117–119
Rational Behavior Inventory, 124
Raven's Progressive Matrices, 58, 61, 135, 191
Recording cognitions, 126
Referral settings, 37–48, 648–671
educational context, 44–46, 648, 658–664
general medical setting, 39–41
legal context, 41–44, 648, 655–657
psychiatric setting, 38–39, 648–655
psychological clinic, 46–48, 649, 664–671

Regression equations, 30, 367
Relationships Beliefs Inventory, 125
Reliability, 10, 12–17
  alternate forms reliability, 14–15
  interscorer reliability, 15–16
  Kuder-Richardson, 13
  split half reliability, 15
  standard error of measurement (SEM), 16–17
  test-retest reliability, 13–14
Renard Diagnostic Interview, 72, 91
Report writing, **621–671**
  case-focused report, 621–622
  content overload, 632
  deciding what to include, 625–627
  emphasis, 627
  feedback, 632–633
  format, 633–647
    background information (history), 637–641
    behavioral observations, 636–639
    behavior and attitude checklist, 637–638
    evaluation procedures, 635–636
    impressions and interpretations, 642–646
    referral question, 634–635
    summary and recommendations, 646–647
    test results, 641–642
  length, 623
  raw data, 629
  sample reports, 647–671
    Case 1: Robert M. (psychiatric setting), 649–655
    Case 2: TOMM evaluation (legal context), 655–657
    Case 3: Anna S. (educational setting), 658–664
    Case 4: A.G. (psychological clinic), 664–671
  shotgun report, 621
  style, 623–624
  terminology, 630–632
  topics, 625
  type of reports, 621–622, 647–649
Research and Diagnostic Criteria (RDC), 91, 93, 96
Resistance, 444–445
Revised Fear Survey Schedule for Children (FSSC-R), 124
Rey Auditory Verbal Learning Test, 522, 527, **558–564**
  administration, 560–562
  alternative lists, 559
  interpretation, 562–564
  norms, 559, 562, 700
  reliability and validity, 559–560
  scoring sheet, 562, 698
Rey-Osterrith Complex Figure Test, 522, 558, 565
Rivermead Behavioural Memory Test, 557
Rogers Criminal Responsibility Scales, 43
Role of the clinician, 3–5
Rorschach, **407–475**, 652–653
  administration, 420–423
  assets and limitations, 416
  blends, 427
  consensus on validity, 419–420
  contents, 428–431
  descriptive statistics, 694–696
  determinants, 424–428
  developmental quality, 424
  history and development, 408–412
  interpretive hypotheses and definitions, 440–475
    AB + Art:R (Intellectualization Index), 436
    Active:Passive ratio (a:p), 436, 466
    Adj D (Adjusted D Score), 436, 465–466
    Adj es (Adjusted es), 436, 465
    Afr (Affective Ratio), 437, 468
    Anatomy and X-Ray, (An + Xy) (Body Concern), 430, 456–457
    animal and animal detail (A and Ad), 429–430, 456
    animal movement (FM), 447–448
    Aspirational Ratio, (W:M), 438, 470–471
    blends (and Blends:R, Complexity Index), 468–469
    Color-achromatic (C′,C′F, FC′), 425, 450–451
    color-chromatic (C, CF, FC, Cn), 425, 449–451
    Complexity index (Blends:R), 468–469
    Conventional Form (X + %), 469
    Conventional Pure Form (F + %), 437, 469
    Coping Deficit Index (CDI), 439, 474
    Depression Index (DEPI), 439, 473
    detail (common; D), 423, 443–444
    detail (unusual; Dd), 423, 444
    Developmental Quality, 424, 445
    Distorted Form (X - %), 469–470
    D Score (D), 436, 465
    eb (Experience Base), 435, 464–465
    Economy Index (W:D:Dd), 438, 470
    Egocentricity index (3r + 2/R), 439, 473
    es (Experienced Stimulation), 435–436
    Experience Actual (EA) (M + C), 435, 463–464
    Experience Balance or Erlebnistypus (EB) (M:C), 435, 462–463
    Experience Pervasiveness (EBPer), 435, 464
    Food, 431, 457
    Form (F), 425, 445–446
    Form Appropriate-Common Areas (WDA%), 437
    Form Appropriate Extended (XA + %), 437
    Form-Color Ratio, FC/(CF + C), 437, 467–468
    Form Dimension (FD), 426, 453
    Form Quality, 427
    human content (H,Hd,(H),(Hd)), 429, 455
    human movement (M), 425, 446–447
    Hypervigilance Index (HVI), 439
    inanimate movement (m), 425, 448–449
    Intellectualization Index, 436, 467
    Interpersonal Interest (Human content), 438, 472
    Isolation Index (Isolate/R), 438–439, 472
    L (Lambda), 435, 445, 461–462

M Active:Passive Ratio, 436, 466–467
Movement, 425, 446–447
Norms, 440
Obsessive Style Index (OBS), 439, 474–475
Organizational Activity (Z), 428, 454
Pairs (2) and Reflections (rF and Fr), 426, 454
Perceptual Thinking Index (PTI), 439, 473
popular responses, 414, 428, 433, 457
Processing Efficiency (Zd), 438, 471–472
responses (number of), 460–461
S - % (White Space Distortion), 470
shading-diffuse (Y,YF,FY), 426, 452–453
shading-dimension (Vista; V,VF,FV), 426, 452
shading-texture (T,TF,FT), 425, 451–452
space (S), 423, 444, 470
special scores, 428–429, 434, 458–460
Suicide Constellation (S-CON), 439, 474
unusual Form (Xu%), 438, 470
White Space Distortion (S - %), 470
whole response, 423, 443
X - % (Distorted Forms), 469–470
location, 423–424
malingering, 416–417
organizational activity, 428
popular responses, 414, 428, 432–433
reliability and validity, 412–416
scoring and interpretive domains, 423–433, 442
special scores, 428–429, 434
structural summary, 433–440
Rorschach Workbook for the Comprehensive System, 423
Rosenzweig Picture Frustration Study, 479
Rotations (Bender), 535
Russell's WMS-R, 198–199

S (Superlative) Scale, 246–247
Sarason Social Support Questionnaire, 606–607
Satz-Mogel/Yudin WAIS/WAIS-R/WAIS-III short form, 194
SCAD/SCALD profiles, 153–154, 187–188
Schedule for Affective Disorders and Schizophrenia (SADS), 77, 91, 93–96
Schizoid, 330
Schizophrenia, indicators of:
MCMI, 346–351, 353
MMPI, 263–265, 281–282, 287, 291–292, 294
neuropsychological tests, 522, 536, 538
Rorschach, 414, 417, 444–446, 450, 453, 455–458, 459, 461, 469, 473
structured interviews, 93–96
Schizotypal personality, 346–348
Scholastic Aptitude Test (SAT), 58
Selecting psychological tests, 62–66
Self efficacy, 125
Self-fulfilling prophecy, 52
Senior Apperception Test, 478
Sensitivity, 22, 97
Sensorium, 88
Sequencing, 151–153, 158, 678–679
Serial thema, 481
Shedler-Westen Assessment Procedure-200 (SWAP-200), 320
Similarities, 163–164, 182, 568–570
Simultaneous processing, 140
Si subscales, 236
Social Avoidance and Distress (SADS), 124
Social Cognition and Object Relations Scale, 481
Social Interaction Self Statement, 124
Social Introversion subscales, 236, 268–270
Social Maturity Index, 405–406
Social skills, 124
Social Support Questionnaire, 607
Somatization, 234, 272–273, 585
Somatoform disorder, 234, 272–273, 585. *See also* Conversion disorder
Southern Mississippi TAT, 479
Specificity, 22, 97
Specific Relationship Standards, 125
Split half reliability, 14–15
Stages of change, 617–619
action, 618–619
contemplation, 618–619
maintenance, 618–619
precontemplation, 618–619
preparation, 618–619
Stages of Change Scale, 618
Stait-Trait Anxiety Inventory (STAI), 318, **590–593**, 604, 616
interpretation, 592–593
reliability and validity, 591–592
Standard error of measurement (SEM), 16, 137, 141
Standardization, 10–12
Standard Progressive Matrices. *See* Raven's Progressive Matrices
Standards for Educational and Psychological Testing, 18, 48
Stanford-Binet, 132, 136, 189
Stanford Shyness Survey, 124
Stroop Test, 522
Structured Clinical Interview for the *DSM-III* (SCID), 96, 646
Structured Interview for *DSM-IV*; Dissociative Disorders (SCID-D), 92
Structured interviews, 91–101
Subjective distress, 41–44, 648, 655–657
Substance abuse, 237, 352, 405, 493
Successive processing, 140
Suicide, indicators of:
BDI, 590
MCMI, 333–334
MMPI, 249–250 276–283
Rorschach, 448, 450, 459, 474
Suicide Constellation (S-CON) (Rorschach), 439, 474
TAT, 500
Supplementary Scales (MMPI-2/MMPI-A), 236, 305
Suppression (on bilateral simultaneous stimulation), 519
Symbol Digit Modalities Test, 556

Symbol Search subtest, 176, 556
Symptom Checklist-90-R (SCL-90-R), 318, **581–587**, 604, 616
  global indexes, 584–585
    Global Severity Index (GSI), 585, 604, 616
    Positive Symptom Distress Index (PSDI), 585
    Positive Symptom Total (PST), 585
  interpretation, 584–587
  norms, 582
  reliability and validity, 582–584
  symptom dimensions, 585–586
    Anxiety (ANX), 586
    Depression (DEP), 585
    Hostility (HOS), 586
    Interpersonal Sensitivity (I-S), 585
    Obsessive-Compulsive (O-C), 585
    Paranoid Ideation (PAR), 586
    Phobic Anxiety (PHOB), 586
    Psychoticism (PSY), 586
    Somatization (SOM), 585
  symptom level/item, 586–587
Systematic Treatment Selection, 581, **599–620**
Systematic Treatment Selection: A Software Program for Treatment Planning, 603
System of Multicultural Pluralistic Assessment (SOMPA), 59–60

Tactical Psychotherapy of the Personality Disorders: An MCMI-III-based Approach, 327
TAT, CAT, and SAT in Clinical Use, 480
Taylor Manifest Anxiety Scale, 613
Tell Me A Story Test (TEMAS), 478–479
Test bias and use with minority groups, 53, 56–62. *See also* Bias
Test of Memory Malingering (TOMM), 655–657
Test-retest reliability, 10, 13–14
Tests: A Comprehensive Reference for Assessment in Psychology, Education, and Business, 10, 62
Tests in Print, II, 10
Thema, 481
Thematic Apperception Test (TAT), **477–515**, 609, 653, 573
  administration, 488–491
  assets and limitations, 485–488
  Bellak TAT and CAT analysis sheet, 506–511, 513
  descriptions of cards, 491–505
  ego function assessment, 508–511
  history and development, 477–480
  interpretation, 514–515
  most useful cards, 488
  Murray's theory of personality, 480–482
  order of presentation, 490
  questioning and inquiry, 490
  reliability and validity, 482–485
  scoring procedures, 505–515
    adequacy of the superego, 511
    conception of the environment, 510
    figures seen as, 510
    integration of the ego, 511
    main defenses against conflicts and fears, 511, 515
    main hero/heroine, 509
    main needs and drives of the hero/heroine, 509
    main theme, 509
    nature of anxieties, 510
    significant conflicts, 510
  theoretical perspectives, 480–482
  typical themes, 491–505
Thematic Test Analysis, 478
Therapeutic Reactance Scale, 613
Therapist Guide to the MMPI & MMPI-2: Providing Feedback and Treatments, 633
Tough Mindedness, 393
Trail Making, 520, 523, 527, **549–556**, 568, 574
  administration, 552–554
  interpretation, 554–556
  norms, 555
  reliability and validity, 552
Treatment planning, 327, 579–593, **595–620**
  BASIC-ID, 600
  Clinician's Rating From, 603
  contraindications for treatment, 605–606
  differential therapeutics, 599
  history and development, 597–601
  MMPI, 238–239. *See also* Minnesota Multiphasic Personality Inventory
  nonspecific features of psychotherapy, 598–599
  pharmacotherapy *vs.* psychotherapy, 605
  prognosis, 605–606
  stages of change (problem solving phase), 602, 617–620
  systematic approach to treatment selection, **599–620**
    coping style, 602, 610–613
    functional impairment, 602, 603–606
    problem complexity/chronicity, 602, 608–610
    resistance (reactance) level, 602, 612–614
    social support, 602, 606–607
    subjective distress, 41–44, 648, 655–657
    therapeutic outcome research, 597–601
    therapist-client matching, 601
True Response Inconsistency (TRIN), 241–242
Two point codes (MMPI-2/MMPI-A), 224–225, 233, 270–294

Universal Nonverbal Intelligence Test, 58, 61, 191
Use of Psychological Testing for Treatment Planning and Outcome Assessment, 239

Validity, 10, 17–31
  conceptual validity, 25
  concurrent validity, 19
  construct validity, 21
  content validity, 18–19
  convergent validity, 21–22
  criterion validity, 19
  ecological (everyday) validity, 139, 199

face validity, 18
incremental validity, 24
predictive validity, 19
sensitivity, 22
specificity, 22
validity in clinical practice, 22–31
Variable Response Inconsistency (VRIN), 241
Vector scale interpretation (CPI), 365, 368–369
Verbal Comprehension Index, 149
Verbal functions and academic skills, 565
Verbal Paired Associates subtest (WMS-III), 210
Verbal-Performance IQ differences, 147–149, 169–170, 179, 180
Vineland Adaptive Behavior Scales, 45, 190
Virtual reality, 66
Visual Delayed Index, 207–208
Visual Immediate Index, 207
Visual-motor perception difficulties, 549
Visuoconstructive abilities, 210
Visuospatial memory, 198
Vocabulary subtest, 162–163, 182, 568

WAIS-III Administration and Scoring Manual, 146, 157, 192
WAIS-III/WMS-III Technical Manual, 201
WAIS-R NI, 564
Wechsler Abbreviated Measure of Intelligence (WASI), 192
Wechsler Adult Intelligence Scale-Third Edition (WAIS-III), 129–195. *See also* Wechsler intelligence scales
Wechsler-Bellevue Intelligence Scale, 132–133, 454
Wechsler Individual Achievement Test (WIAT), 45, 135–136, 183, 189, 566, 660–661
Wechsler intelligence scales (WAIS/WAIS-R/WAIS-III, WISC/WISC-R/WISC-III/WISC-IV), **129–195**, 519, 523, 527, 652, 660, 663, 677–689
ACID/SCAD profiles, 146, 151, 153–154, 186–188
administration time, 144
administrative checklist, 144
Alzheimer's disease, 185–186
assets and limitations, 137–140
Bannatyne's categories, 146, 151, 152–153, 574, 678–679
brain damage, 170–173, 178–182
cautions and guidelines for administration, 142–145
classifications of intelligence, 142–143
clerical errors, 139, 141
deviation IQ, 141
errors in administration, 139
estimating premorbid level of functioning, 182–185
ethnic minorities, 139, 190–191
factors (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, Processing Speed), 135–137
Fuld profile, 185–186
Full Scale IQ, 147
gifted children, 190
history and development, 132
Horn groupings, 1i1, 154–155, 678–679
hypothesis testing for subtests, 682–689
indexes (WISC-III and WAIS-III), 149–153
interpretation procedures, 145–160, 678–689
intersubtest variability, 146, 155–159, 179–180, 678–689
intrasubtest variability, 159–160
learning disabilities, 186–188, 678–679
meaning of IQ scores, 140
mental retardation, 188–190
percentile rankings for Wechsler deviation IQs, 142, 677
Perceptual Organization Index, 149–150
practice effects, 134
Processing Speed Index, 150–151, 179–180
profile analysis, 146, 155–159, 179–180, 678–689
qualitative analysis, 160, 182
reliability and validity, 134–137
SCAD/SCALD profiles, 146, 151, 153–154, 187–188
short forms, 191–195
standard error of measurement (SEM), 137, 141
subtest variability within scales, 139–139, 159–160
Verbal Comprehension Index, 149
Verbal IQ, 148–149
Verbal/Performance differences, 146–149, 152–162, 169–170, 179–181
Wechsler subtests, 160–178
arithmetic, 164, 181, 568–569
block design, 172, 181, 532, 547–548
comprehension, 167–168, 568–569
digit span, 165, 181, 556, 564, 568–569
digit symbol (coding), 170–172, 181, 532, 556, 568
information, 166–167, 182, 569
letter-number sequencing, 168, 564, 568–569
matrix reasoning, 174
mazes, 177–178
object assembly, 176, 547–548
picture arrangement, 174–176, 181
picture completion, 170, 182, 566
similarities, 163–164, 182, 568–569
symbol search, 176, 556
vocabulary, 162–163, 182, 569–570
Working Memory Index, 150, 165
Wechsler Memory Scales (WMS/WMS-R/WMS-III), 135, 138, **197–212**, 523, 557, 565
assets and limitations, 203–204
factor analysis, 199, 202
history and development, 198–201
indexes, 200, 204–212
Auditory Delayed/Auditory Recognition Delayed, 208
Auditory Delayed/Visual Delayed, 208
Auditory Immediate/Visual Immediate, 207
General Memory Index, 204, 206, 209

Wechsler Memory Scales (WMS/WMS-R/WMS-III) *(Continued)*
Immediate Memory/General Memory (Delayed), 206
Visual Immediate/Visual Delayed, 207
Working Memory/General Memory, 209
Working Memory/Immediate Memory, 208
interpretation, 204–212
attention, 210
attention versus acquisition/encoding, 208–209
auditory versus visual modalities, 207–208
change, 211
immediate versus delayed memory, 206–207
information and orientation, 210
IQ versus General Memory differences, 204–205
Logical Memory I and II, 211
learning, 210
Letter-Number Sequencing, 211
orientation, 209
recall versus recognition, 208
Verbal paired associates, 210
Visual construction, 210
IQ-Memory Quotient differences, 204–205
Memory Quotient, 198–199
norms, 199–203
reliability and validity, 199–200, 201–202
Russell's WMS-R, 198–199
short forms, 203
standard error of measurement, 199
standardization, 199–203
WAIS-III/WMS-III Technical Manual, 201
WMS, 198
WMS-III Administration and Scoring Manual, 204, 206
WMS-R, 199–200
Wechsler Preschool and Primary Scale of Intelligence (WPPSI/WPPSI-R/WPSSI-III), 133
Wechsler Test of Adult Reading, 183
Welsh Anxiety Scale, 590–592
Wide Range Achievement Test-III (WRAT-III), 8, 45, 183, 566
Wide Range Assessment of Memory and Learning, 557
Wiggins Content Scales, 218
WISC-III Manual, 146, 157
WISC-III PI, 564
WMS-III Administration and Scoring Manual, 204, 206
Wolpe-Lazarus Assertion Inventory, 124
Woodcock-Johnson Psychoeducational Battery-III, 44
Woodworth Personal Data Sheet, 581
Word finding difficulties, 518–519, 565–566
Working memory, 197, 208–209
Working Memory Index, 150, 165, 208–209